UNIVERSITY OF ALBERTA

# Robust Designs for Approximate Regression Models with Two Interacting Regressors

by

Adeniyi Johnson Adewale

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the

requirements for the degree of Master of Science

in

**Statistics**

Department of Mathematical and Statistical Sciences

Edmonton, Alberta

Spring 2003

National Library of Canada
Bibliothèque nationale du Canada

Acquisitions and Bibliographic Services
Acquisitions et services bibliographiques

395 Wellington Street
Ottawa ON K1A 0N4
Canada

395, rue Wellington
Ottawa ON K1A 0N4
Canada

*Your file Votre référence*

*Our file Notre référence*

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-82241-9

Canada

**UNIVERSITY OF ALBERTA**

**Library Release Form**

**Name of Author:** *Adeniyi Johnson Adewale*

**Title of Thesis:** *Robust Designs for Approximate Regression Models with Two Interacting Regressors*

**Degree:** *Master of Science*

**Year this Degree Granted:** *2003*

Permission is hereby granted to the University of Alberta Library to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly or scientific research purposes only.

The author reserves all other publication and other rights in association with the copyright in the thesis, and except as herein before provided, neither the thesis nor any substantial portion thereof may be printed or otherwise reproduced in any material form whatever without the author's prior written permission.

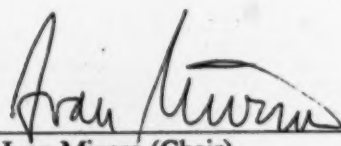
*Department of Mathematical & Statistical Sciences*
*632, Central Academic Building*
*University of Alberta*
*Edmonton, Alberta*

*Date:* DECEMBER 18, 2002

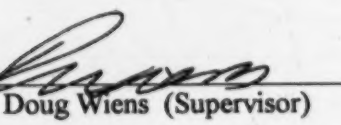
UNIVERSITY OF ALBERTA

Faculty of Graduate Studies and Research

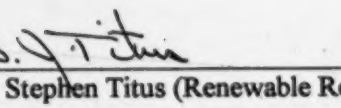
The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled **Robust Designs For Approximate Regression Models With Two Interacting Regressors** submitted by **Adeniyi Johnson Adewale** in partial fulfillment of the requirements for the degree of **Master of Science** in Statistics.



Dr. Ivan Mizera (Chair)



Dr. Doug Wiens (Supervisor)



Dr. Stephen Titus (Renewable Resources)

December 9, 2002

**To God who alone is wise**

## ABSTRACT

In this thesis we set out to investigate the approximately linear regression model $E(Y|\mathbf{x}) = \mathbf{z}'(\mathbf{x})\boldsymbol{\theta}$ given that the true model is $E(Y|\mathbf{x}) = \mathbf{z}'(\mathbf{x})\boldsymbol{\theta} + f(\mathbf{x})$, where $f(\mathbf{x})$ is unknown but restricted by a bound on its $\mathcal{L}_2(\mathcal{S})$ norm and $\mathcal{S}$ is the design space. We consider the case where $\mathbf{z}'(\mathbf{x})$ has interaction terms, in particular $\mathbf{x} = (x_1, x_2)'$, $\mathbf{z}'(\mathbf{x}) = (1, x_1, x_2, x_1x_2)$. Goodness of fit of the approximation given by $E(Y|\mathbf{x}) = \mathbf{z}'(\mathbf{x})\boldsymbol{\theta}$ is measured using functions of the Mean Squared Error matrix, $MSE(\mathbf{x}) = E[\{z'(\mathbf{x})\hat{\boldsymbol{\theta}}_0 - E(Y|\mathbf{x})\}^2]$. In particular, we examine the Integrated Mean Squared Error, Determinant of Mean Squared Error and Trace of Mean Squared Error as overall measures of loss. Using the minimax approach we construct optimal designs which are robust against model misspecification. We employ Lagrange multipliers to determine the form of the optimal density corresponding to each measure of overall loss and using numerical methods we derive explicit expressions for the optimal densities. Finally, we illustrate the implementations of the optimal designs.

## ACKNOWLEDGEMENTS

I would like to express my sincere appreciation to Dr. Douglas P. Wiens for granting me the privilege to write this thesis under his supervision and for his support, both financial and otherwise, through out the course of my program. This work would have been impossible without his guidance and help. Special thanks to members of my examining committee Dr. Ivan Mizera and Dr. Stephen Titus for their helpful comments and suggestions on this work.

I thank my professors, Dr. K.C Carriere, Dr. P.M. Hooper and Dr. R.J. Karunamuni who have shared with me their wealth of statistical knowledge at various points in the course of my program.

My gratitude to Mr and Mrs Adebimpe Adewale for their parental support. I do not forget to acknowledge the unwavering love of my fiancee, Lola Odumakinde. Many thanks to friends and brethren of like precious faith whose encouragements and prayers have been instrumental to the completion of this work.

# Contents


**1 LITERATURE REVIEW AND OBJECTIVE** **1**

1.1 Introduction . . . . . 1

1.2 Some Earlier Work in Robust Designs . . . . . 4

**2 MINIMAX DESIGNS** **7**

2.1 Introduction . . . . . 7

2.2 Integrated Mean Squared Error . . . . . 10

2.3 Other Overall Measures of Loss . . . . . 18

2.4 Minimizing Maximum Loss over the Designs $\xi$ . . . . . 20

**3 ANALYTICAL AND NUMERICAL RESULTS** **25**

3.1 Optimal Density Corresponding to First Eigenvalues . . . . . 26

3.2 Optimal Density Corresponding to Second Eigenvalues . . . . 28

3.3 Optimal Density Corresponding to Third Eigenvalues . . . . . 30

3.4 Optimal Coefficients for small $\nu$ . . . . . 31

3.5 Larger Values of $\nu$ for Q- and D-optimality . . . . . 37

3.6 Optimal Coefficients for A-Optimality . . . . . 45

**4 IMPLEMENTATIONS** **51**

4.1 Case Study 1: Waste Water Ozonation . . . . . 51

4.2 Case Study 2: Injection Molding Process . . . . . 53

**REFERENCES** **55**

**APPENDIX: SAMPLE PROGRAM CODES** **57**

# Chapter 1

# LITERATURE REVIEW AND OBJECTIVE

## 1.1 Introduction

Suppose that an experimenter fits, by least squares, a regression model

$$E(Y|\mathbf{x}) = \mathbf{z}'(\mathbf{x})\theta \tag{1.1}$$

to data $\{(Y_i, \mathbf{x}_i)\}_{i=1}^n$, with the $\mathbf{x}_i$ being chosen from a $q$-dimensional design space $\mathcal{S}$. The mean response is linear in $p$ regressors $z_1(\mathbf{x}), ..., z_p(\mathbf{x})$, each a function of independent variables $x_1, ..., x_q$.

Given the data $\{(Y_i, \mathbf{x}_i)\}_{i=1}^n$, the classical regression model is given by

$$y(\mathbf{x}_i) = z'(\mathbf{x}_i)\theta + \varepsilon_i, i = 1, 2, ..., n \tag{1.2}$$

where the $\varepsilon_i$ are uncontrollable random errors. Two classical assumptions about the model (1.2) are

A1. The regression response $E(Y|\mathbf{x}) = z'(\mathbf{x})\theta$ is exactly correct.

A2. The errors $\varepsilon_i$ are uncorrelated and have variance $\sigma^2$.

Define $\mathbf{X} = (z(x_1), ..., z(x_n))'$ and $Y = (y_1, ..., y_n)'$. Under the assumptions A1 and A2, the Least Squares estimate $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ is given by

$$\hat{\boldsymbol{\theta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'Y.$$

The Least Squares estimate $\hat{\boldsymbol{\theta}}$ is the best linear unbiased estimate of $\boldsymbol{\theta}$. We know $\hat{\boldsymbol{\theta}}$ has covariance matrix

$$cov(\hat{\boldsymbol{\theta}}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1},$$

which only depends on $\mathbf{X}$, the so-called "model matrix". Hence, the problem of experimental design is to choose the appropriate $\mathbf{X}$ in order that $cov(\hat{\boldsymbol{\theta}})$ be as small as possible.

For example, suppose an experimenter is interested in measuring water purity ($y$), as a function of chlorine added ($x$), assuming $x$ has been coded such that $x \in [-1, 1]$. When observations ($y$) are subject to experimental error, the $n$ observations are given by

$$y(x_i) = \theta_0 + \theta_1 x_i + \varepsilon_i, \; i = 1, ..., n, \; -1 \leq x_i \leq 1.$$

where $x_i$ is the quantity of chlorine added and $y(x_i)$ is the level of purity of water corresponding to the quantity $x_i$. We assume the error terms are uncorrelated with mean zero and variance $\sigma^2$. The parameters $\theta_0$ and $\theta_1$ are unknown regression coefficients. Using least squares, the estimates of $\theta_0$ and $\theta_1$ are given by,

$$\begin{aligned} \hat{\theta}_1 &= \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2}, \\ \hat{\theta}_0 &= \bar{y} - \hat{\theta}_1 \bar{x}. \end{aligned}$$

In terms of $s_x^2 = \sum(x_i - \bar{x})^2/n$ one finds that

$$nVAR(\hat{\theta}_1) = \sigma_\varepsilon^2/s_x^2 \quad (1.3)$$

$$nVAR(\hat{\theta}_0) = \sigma_\varepsilon^2(1 + \bar{x}^2/s_x^2). \quad (1.4)$$

Classical design theory stipulates that the optimal design is the design that minimizes the variances of our estimates. That is, the design that minimizes the expressions in (1.3), (1.4). This is achieved by putting 1/2 of the $x's$ at each of $\pm 1$, so that $\bar{x}^2 = 0$ (minimum) and $s_x^2 = 1$ (maximum). In our terminology the design measure $\xi(x)$ (the fraction of design points placed at $x$) is given by

$$\xi(x) = \begin{cases} 1/2, & x = \pm 1 \\ 0, & \text{otherwise.} \end{cases}$$

Suppose the experimenter had wanted to fit the quadratic model

$$y(x_i) = \theta_0 + \theta_1 x_i + \theta_2 x_i^2 + \varepsilon_i,$$
$$i = 1, ..., n, -1 \leq x_i \leq 1,$$

where $x_i$ remains the quantity of chlorine added and $y(x_i)$ the level of purity of water corresponding to $x_i$. Here the classical optimal design measure, minimizing the determinant of $cov(\hat{\boldsymbol{\theta}}) = \sigma_\varepsilon^2(\mathbf{X}'\mathbf{X})^{-1}$ is given by

$$\xi(x) = \begin{cases} 1/3, & x = 0, \pm 1 \\ 0, & \text{otherwise.} \end{cases}$$

If the experimenter is mistaken about the choice of the relationship between water purity and quantity of chlorine added, the classical optimal designs for both models considered cannot possibly detect the presence of

departures from the prescribed models. A major drawback of classical optimal designs is their extreme model dependency. These designs provide no opportunity to check the model's adequacy. If the mean response (1.1) is not exactly linear in the $p$ regressors $z_1(\mathbf{x}), ..., z_p(\mathbf{x})$, then the optimal design cannot possibly detect the presence of any departure from linearity in the regressors $z_1(\mathbf{x}), ..., z_p(\mathbf{x})$, no matter how large the sample size.

## 1.2 Some Earlier Work in Robust Designs

Box and Draper (1959) in the paper, *A basis for the selection of a response surface design*, revealed the inherent dangers of designing a regression experiment on the basis that (1.1) is exactly correct. They studied the case where the experimenter fits a polynomial of first degree whereas the correct model is quadratic. The estimate is subject to "bias error" engendered by the model misspecification as well as "variance error" due to sampling. Box and Draper (1959) concluded that any slight model misspecification can erode any suppose gains arising from the use of a design which minimize variance alone.

Subsequent to Box and Draper (1959), various authors have investigated realistic designs, in which case (1.1) is taken as an approximation of the true model, a more precise description being

$$E(Y|\mathbf{x}) = \mathbf{z}'(\mathbf{x})\boldsymbol{\theta} + f(\mathbf{x}) \tag{1.5}$$

for some unknown but "small" function $f$. Some authors like Marcus and Sacks (1976), Li and Notz (1982), Pesotchinsky (1982), Sacks and Ylvisaker (1978) and Liu and Wiens (1997) have discussed designs in which $f(\mathbf{x})$ belongs to the class

$$\mathcal{F} = \{f : |f(\mathbf{x})| \leq \phi(\mathbf{x}), \forall \mathbf{x} \in \mathbf{S}\}, \tag{1.6}$$

with various assumptions being made about $\phi$.

Huber (1975) takes $f(\mathbf{x})$ from

$$\mathcal{F} = \{f\colon \int_S \mathbf{z}(\mathbf{x})f(\mathbf{x})\,d\mathbf{x} = 0, \int_S f^2(\mathbf{x})\,d\mathbf{x} \leq \eta^2\}. \tag{1.7}$$

The radius $\eta$ of $\mathcal{F}$ is assumed fixed. The first condition in $\mathcal{F}$ says that $f$ and $\mathbf{z}$ are orthogonal, so the parameter $\theta$ is uniquely defined in model (1.5). To see this suppose that

$$E(Y|\mathbf{x}) = \mathbf{z}'(\mathbf{x})\boldsymbol{\theta}_1 + f_1(\mathbf{x})$$

and that

$$E(Y|\mathbf{x}) = \mathbf{z}'(\mathbf{x})\boldsymbol{\theta}_2 + f_2(\mathbf{x}),$$

where $\mathbf{z}(\mathbf{x})$ is orthogonal to $f_1$and $f_2$. Thus

$$\begin{aligned} 0 &= \int_S \mathbf{z}(\mathbf{x})(f_2(\mathbf{x}) - f_1(\mathbf{x}))d\mathbf{x} \\ &= \int_S \mathbf{z}(\mathbf{x})\mathbf{z}'(\mathbf{x})\,d\mathbf{x}(\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2) \end{aligned}$$

so that, under the assumption that $\int_S \mathbf{z}(\mathbf{x})\mathbf{z}'(\mathbf{x})\,d\mathbf{x}$ is positive definite, we have that $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_2$. Thus $\boldsymbol{\theta}$ is unique under the first condition of $\mathcal{F}$. The second condition assumes that $f$ is small otherwise the model (1.5) will be totally wrong.

Some controversy revolves around the choice of the class $\mathcal{F}$. Marcus and Sacks (1976) and Li and Notz (1982) have criticized the class (1.7) as being too full. Wiens (1992) reported that the class (1.6) is thin, however, since it seems invariably to lead to 'robust' designs, all of whose mass is concentrated at a small number of, generally extreme, points in the design space. Wiens

(1992) posited that an approximation to a design which is robust against more realistic alternatives is preferable to an exact solution in a neighbourhood which is unrealistically sparse.

Heo, Schmuland and Wiens (2001) also considered designs in (1.7) (Huber's class). Their design depend on $\eta^2$ and the common variance, $\sigma^2$ of random errors only through $\nu = \sigma^2/(n\eta^2)$, which may be chosen by the experimenter according to her judgement of the relative importance of variance versus bias. Wiens (1990) investigated the model,

$$Y = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_{12} x_1 x_2 + f(\mathbf{x}) + \varepsilon, 1/2 \leq x_1, x_2 \leq 1/2,$$

with two interacting regressors. Solutions were given only for small values of $\nu$ using Integrated Mean Squared Error loss.

In this thesis work we seek to proffer solutions for larger values of $\nu$. In addition we shall consider other overall measures of loss apart from Integrated Mean Squared Error.

# Chapter 2

# MINIMAX DESIGNS

## 2.1 Introduction

In this thesis, we aim to derive minimax design for regression models. The minimax approach minimizes (over a class of designs) the maximum (over $\mathcal{F}$) value of a mean of the "loss" - here, the mean squared error matrix.

With the knowledge that the true model might be only approximated by (1.1) and that a more precise model is given by (1.5), we seek to choose design points that yield estimates $\hat{\theta}$ of $\theta$ and estimates $\hat{Y}(\mathbf{x}) = z'(\mathbf{x})\hat{\theta}$ of $E(Y|\mathbf{x})$ which remain relatively efficient while suffering as little as possible from the bias engendered by the model misspecification.

We measure the goodness of fit of the approximation given by (1) by Mean Squared Error,

$$MSE(\mathbf{x}) = E[\{z'(\mathbf{x})\hat{\boldsymbol{\theta}}_0 - E(Y|\mathbf{x})\}^2]$$

that is, the average (squared) amount by which our estimate $z'(\mathbf{x})\theta$ "misses" the true mean response $E(Y|\mathbf{x})$.

Heo, Schmuland and Wiens (2001) reported three kind of loss functions, $\mathcal{L}_Q, \mathcal{L}_D, \mathcal{L}_A$ which represented the integrated MSE of the fitted response $\hat{Y}(\mathbf{x})$, determinant of the MSE matrix and the trace of the MSE matrix respectively. In this thesis we shall construct optimal designs using each of these three loss functions.

Since (1.1) is just an approximation of the true model, the "best" $\theta_0$, for predicting the mean response $E(Y|\mathbf{x})$, is defined by

$$\theta_0 = argmin \int_S \{z'(\mathbf{x})\theta - E(Y|\mathbf{x})\}^2 \, d\mathbf{x}.$$

Define

$$f(\mathbf{x}) = \mathbf{z}'(\mathbf{x})\boldsymbol{\theta}_0 - E(Y|\mathbf{x})$$

then our approximate linear model is

$$Y = \mathbf{z}'(\mathbf{x})\boldsymbol{\theta}_0 + f(\mathbf{x}) + \epsilon, \mathbf{x} \in S, \tag{2.1}$$

Under (2.1), the parameter $\boldsymbol{\theta}$ is well-defined if $f$ is constrained by (1.7). Our model then becomes

$$Y(x_i) = E(Y|x_i) + \varepsilon_i, i = 1, ..., n$$

with the mean response $E(Y|\mathbf{x}) = \mathbf{z}'(\mathbf{x})\boldsymbol{\theta} + f(\mathbf{x})$ and with $f$ an arbitrary, unknown member of

$$\mathcal{F} = \{f: \int_S \mathbf{z}(\mathbf{x}) f(\mathbf{x}) \, d\mathbf{x} = 0, \int_S f^2(\mathbf{x}) \, d\mathbf{x} \leq \eta^2\}.$$

The mean squared error is

$$\begin{aligned} MSE &= E[(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})'] \\ &= E[(\hat{\boldsymbol{\theta}} - E(\hat{\boldsymbol{\theta}}))(\hat{\boldsymbol{\theta}} - E(\hat{\boldsymbol{\theta}}))'] + (E(\hat{\boldsymbol{\theta}}) - \boldsymbol{\theta})(E(\hat{\boldsymbol{\theta}}) - \boldsymbol{\theta})' \\ &= cov(\hat{\boldsymbol{\theta}}) + \mathbf{bias}.\mathbf{bias}' \end{aligned}$$

An exactly implementable design will correspond to a design measure $\xi$ placing mass $n^{-1}$ at each of $x_1, ..., x_n$. This design is what we call the empirical design measure.

We define the probability mass function of the empirical measure as

$$P_\xi(x = x_i) = \frac{1}{n},\ i = 1, ..., n.$$

Then,

$$\begin{aligned} cov(\hat{\theta}) &= \sigma_\varepsilon^2(\mathbf{X}'\mathbf{X})^{-1} \\ &= \sigma_\varepsilon^2\left(\sum_{i=1}^{n} z(x_i)z'(x_i)\right)^{-1} \\ &= \frac{\sigma_\varepsilon^2}{n}\left[\frac{1}{n}\sum_{i=1}^{n} z(x_i)z'(x_i)\right]^{-1} \\ &= \frac{\sigma_\varepsilon^2}{n}\left(E_\xi[\mathbf{z}(\mathbf{x})\mathbf{z}'(\mathbf{x})]\right)^{-1} \\ &= \frac{\sigma_\varepsilon^2}{n}\mathbf{B}^{-1}(\boldsymbol{\xi}) \end{aligned}$$

where $\mathbf{B}(\xi) = E_\xi[\mathbf{z}(\mathbf{x})\mathbf{z}'(\mathbf{x})]$ depends only on the design $\xi$, not on $f$.

The bias is given by

$$\mathbf{bias} = E(\hat{\theta} - \theta)$$

$=\mathrm{E}[(\mathrm{X}'\mathbf{X})^{-1}(\mathbf{X}'Y)] - \theta$

$=\mathrm{E}[(\mathrm{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\boldsymbol{\theta} + f + \varepsilon)] - \theta$

$=(\mathrm{X}'\mathbf{X})^{-1}\sum z(x_i)f(x_i)$

$=(\mathrm{nB}(\xi))^{-1}nE_\xi[\mathbf{z}(\mathbf{x})f(\mathbf{x})]$

$=\mathrm{B}^{-1}(\xi)\mathbf{b}(f,\xi)$where $\mathbf{b}(f,\xi) = E_\xi[\mathbf{z}(\mathbf{x})f(\mathbf{x})]$. Thus

$$MSE(f,\xi) = \frac{\sigma_\varepsilon^2}{n}\mathbf{B}^{-1}(\xi) + \mathbf{B}^{-1}(\xi)\mathbf{b}(f,\xi)\mathbf{b}'(f,\xi)\mathbf{B}^{-1}(\xi). \qquad (2.2)$$

## 2.2 Integrated Mean Squared Error

In this section we shall construct our overall measure of the loss based on the "Integrated Mean Squared Error". We define the Integrated Mean Squared Error as

$$IMSE = \int_S MSE(\mathbf{x})\, d\mathbf{x}$$

where $MSE(\mathbf{x})$ is the mean squared error of $z'(\mathbf{x})\hat{\theta}$ given by

$$\begin{aligned} MSE(\mathbf{x}) &= E[(\mathbf{z}'(\mathbf{x})\hat{\boldsymbol{\theta}} - E(Y|\mathbf{x}))^2] \\ &= E[(\mathbf{z}'(\mathbf{x})\hat{\boldsymbol{\theta}} - \mathbf{z}(\mathbf{x})\boldsymbol{\theta} - f(\mathbf{x}))^2] \\ &= E[(\mathbf{z}'(\mathbf{x})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}))^2] - 2f(\mathbf{x})E[\mathbf{z}'(\mathbf{x})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})] + f^2(\mathbf{x}) \\ &= \mathbf{z}'(\mathbf{x})E[(\hat{\boldsymbol{\theta}} - \theta)(\hat{\boldsymbol{\theta}} - \theta)']\mathbf{z}(\mathbf{x}) - 2f(\mathbf{x})\mathbf{z}'(\mathbf{x})E(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \\ &\quad + f^2(\mathbf{x}) \end{aligned}$$

Thus the "Integrated Mean Squared Error" is given by

$$\begin{aligned} IMSE &= \int_S MSE(\mathbf{x})\, d\mathbf{x} \\ &= \int_S \{\mathbf{z}'(\mathbf{x})MSE(f,\xi)\mathbf{z}(\mathbf{x})\, d\mathbf{x} - 2f(\mathbf{x})\mathbf{z}'(\mathbf{x})E(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \\ &\quad + f^2(\mathbf{x})\}\, d\mathbf{x}. \end{aligned}$$

Using $\int_S \mathbf{z}(\mathbf{x})f(\mathbf{x})\, d\mathbf{x} = \mathbf{0}$, we have

$$IMSE = \int_S \mathbf{z}'(\mathbf{x})MSE(f,\xi)\mathbf{z}(\mathbf{x})\, d\mathbf{x} + \int_S f^2(\mathbf{x})\, d\mathbf{x}.$$

But,

$$\begin{aligned}\int_S \mathbf{z}'(\mathbf{x})MSE(f,\xi)\mathbf{z}(\mathbf{x})\,d\mathbf{x} &= \int_S tr\{\mathbf{z}'(\mathbf{x})MSE(f,\xi)\mathbf{z}(\mathbf{x})\}\,d\mathbf{x} \\ &= \int_S tr\{MSE(f,\xi)\cdot\mathbf{z}(\mathbf{x})\mathbf{z}'(\mathbf{x})\,d\mathbf{x} \\ &= tr\{MSE(f,\xi)\cdot\int_S \mathbf{z}(\mathbf{x})\mathbf{z}'(\mathbf{x})\,d\mathbf{x}\} \\ &= tr\{MSE(f,\xi)\cdot\mathbf{A}\},\end{aligned}$$

where $\mathbf{A} = \int_S \mathbf{z}(\mathbf{x})\mathbf{z}'(\mathbf{x})\,d\mathbf{x}$.

Using (2.2), we have

$$\begin{aligned}& tr\{MSE(f,\xi)\cdot\mathbf{A}\} \\ =\ & tr\{\frac{\sigma_\varepsilon^2}{n}\mathbf{B}^{-1}(\xi)\cdot\mathbf{A}\} + tr\{\mathbf{B}^{-1}(\xi)\mathbf{b}(f,\xi)\mathbf{b}'(f,\xi)\mathbf{B}^{-1}(\xi)\cdot\mathbf{A}\} \\ =\ & \frac{\sigma_\varepsilon^2}{n}tr\{\mathbf{B}^{-1}(\xi)\cdot\mathbf{A}\} + tr\{\mathbf{b}'(f,\xi)\mathbf{B}^{-1}(\xi)\cdot\mathbf{A}\cdot\mathbf{B}^{-1}(\xi)\mathbf{b}(f,\xi)\} \\ =\ & \frac{\sigma_\varepsilon^2}{n}tr\{\mathbf{B}^{-1}(\xi)\cdot\mathbf{A}\} + \mathbf{b}'(f,\xi)\mathbf{H}^{-1}(\xi)\mathbf{b}(f,\xi),\end{aligned}$$

where $\mathbf{H}(\xi) = \mathbf{B}(\xi)\mathbf{A}^{-1}\mathbf{B}(\xi)$. Thus,

$$IMSE = \frac{\sigma_\varepsilon^2}{n}tr\{\mathbf{B}^{-1}(\xi)\cdot\mathbf{A}\} + \mathbf{b}'(f,\xi)\mathbf{H}^{-1}(\xi)\mathbf{b}(f,\xi) + \int_S f^2(\mathbf{x})\,d\mathbf{x}.$$

The loss function is given by

$$\begin{aligned}& max_{\mathcal{F}}\mathcal{L}_Q(f,\xi) = max_{\mathcal{F}}IMSE \\ =\ & \frac{\sigma_\varepsilon^2}{n}tr\{\mathbf{B}^{-1}(\xi)\cdot\mathbf{A}\} + max_{\mathcal{F}}\{\mathbf{b}'(f,\xi)\mathbf{H}^{-1}(\xi)\mathbf{b}(f,\xi) \\ & + \int_S f^2(\mathbf{x})\,d\mathbf{x}\}\end{aligned}$$

Thus we seek to maximize IMSE over the class $\mathcal{F}$. We claim that IMSE attains its maximum at an $f(\mathbf{x})$ with $\int_S f^2(\mathbf{x})\,d\mathbf{x} = \eta^2$. To see this, suppose

$\int_S f^2(\mathbf{x})\,d\mathbf{x} = c^2 < \eta^2$. Define

$$f_c(\mathbf{x}) = \frac{\eta f(\mathbf{x})}{c}.$$

Clearly,

$$\int_S \mathbf{z}(\mathbf{x}) f_c(\mathbf{x}) = 0$$

and

$$\int_S f_c^2(\mathbf{x})\,d\mathbf{x} = \eta^2.$$

Thus $f_c(\mathbf{x}) \in \mathcal{F}$ and $c^2 = \int_S f^2(\mathbf{x})\,d\mathbf{x} < \int_S^2 f_c(\mathbf{x})\,d\mathbf{x} = \eta^2$.

Now,

$$\begin{aligned}
\mathbf{b}(f_c, \xi) &= \int_S \mathbf{z}(\mathbf{x}) f_c(\mathbf{x}) d\xi(\mathbf{x}) \\
&= \frac{\eta}{c} \int_S \mathbf{z}(\mathbf{x}) f(\mathbf{x}) d\xi(\mathbf{x}) \\
&= \frac{\eta}{c} \mathbf{b}(f, \xi).
\end{aligned}$$

Thus

$$\begin{aligned}
\mathbf{b}'(f_c, \xi)\mathbf{H}^{-1}(\xi)\mathbf{b}(f_c, \xi) &= \frac{\eta^2}{c^2} \mathbf{b}'(f, \xi)\mathbf{H}^{-1}(\xi)\mathbf{b}(f, \xi) \\
&> \mathbf{b}'(f, \xi)\mathbf{H}^{-1}(\xi)\mathbf{b}(f, \xi)
\end{aligned}$$

since

$$\frac{\eta^2}{c^2} > 1.$$

Therefore, we have established that for any $f \in \mathcal{F}$ such that $\int_S f^2(\mathbf{x})\,d\mathbf{x} < \eta^2$, there exist $f_c \in \mathcal{F}$ which results in a larger IMSE. Hence, the maximum IMSE is attained at an $f$ with $\int_S f^2(\mathbf{x})\,d\mathbf{x} = \eta^2$.

Thus,

$$max_{\mathcal{F}}\mathcal{L}_Q(f,\xi) = \frac{\sigma_\epsilon^2}{n} tr\{\mathbf{B}^{-1}(\xi)\cdot\mathbf{A}\} + max_{\mathcal{F}_1}\{\mathbf{b}'(f,\xi)\mathbf{H}^{-1}(\xi)\mathbf{b}(f,\xi)\} + \eta^2$$

where

$$\mathcal{F}_1 = \{f\colon \int_S \mathbf{z}(\mathbf{x})f(\mathbf{x})\,d\mathbf{x} = 0, \int_S f^2(\mathbf{x})\,d\mathbf{x} = \eta^2\}.$$

So far, in our derivations, we have used the empirical design measure with probability mass function $P_\xi(x = x_i) = \frac{1}{n}, i = 1, ..., n$. In practice, $\xi(x_i)$ must be a discrete probability measure, with $\xi(x_i), i = 1, 2, ..., n$, being an integral multiple of $\frac{1}{n}$. A design with this integral property is called an exact design. But the exact design problem is very difficult and mathematically intractable. In addition to the problem of intractability Wiens (1992) and Heo, Schmuland and Wiens (1999) proved that in order for $max_{\mathcal{F}}\mathcal{L}_Q(f,\xi)$ to be finite, it is necessary that $\xi$ be absolutely continuous with respect to Lebesque measure, with a density $m(\mathbf{x})$ satisfying $\int_S m^2(\mathbf{x})\,d\mathbf{x} < \infty$. Henceforth, we assume that $\xi$ is absolutely continuous with respect to Lebesque measure.

Now,

$$\mathbf{B}(\xi) = \int_S \mathbf{z}(\mathbf{x})\mathbf{z}'(\mathbf{x})m(\mathbf{x})\,d\mathbf{x}, \tag{2.3}$$

$$\mathbf{b}(f,\xi) = \int_S \mathbf{z}(\mathbf{x})f(\mathbf{x})m(\mathbf{x})\,d\mathbf{x}. \tag{2.4}$$

Define

$$\mathbf{K}(\xi) = \int_S \mathbf{z}(\mathbf{x})\mathbf{z}'(\mathbf{x})m^2(\mathbf{x})\,d\mathbf{x}. \tag{2.5}$$

To conclude this maximization we are left to maximize

$$B(f,\xi) = \mathbf{b}'(f,\xi)\mathbf{H}^{-1}(\xi)\mathbf{b}(f,\xi)$$

subject to

$$\begin{aligned}\int_S \mathbf{z}(\mathbf{x})f(\mathbf{x})\,d\mathbf{x} &= 0,\\ \int_S f^2(\mathbf{x})\,d\mathbf{x} &= \eta^2.\end{aligned}$$

The derivation that follows is as in Wiens (1990), and is included for completeness. Let $\mathcal{F}_2$ be the convex set defined by

$$\mathcal{F}_2 = \{f\colon \int_S f^2(\mathbf{x})\,d\mathbf{x} \le \eta^2\},$$

and for $f_0, f_1 \in \mathcal{F}_2$ and $\lambda \in [0,1]$ put $f_\lambda = (1-\lambda)f_0 + \lambda f_1$. Let $\mathbf{d} \in \mathcal{R}^p, d_0 \in \mathcal{R}$ be Lagrange multipliers, and put

$$\phi(\lambda) = B(f_\lambda,\xi) + \mathbf{d}'\int_S \mathbf{z}(\mathbf{x})f_\lambda(\mathbf{x})\,d\mathbf{x} + d_0\int_S f_\lambda^2(\mathbf{x})\,d\mathbf{x}.$$

A maximizing $f_0$ must then satisfy, for each $f_1$,

$$\begin{aligned}0 &\ge \phi'(0)\\ &= \int_S (f_1-f_2)(\mathbf{x})[2\mathbf{b}'(f_0,\xi)\mathbf{H}^{-1}(\xi)\mathbf{z}(\mathbf{x})m(\mathbf{x}) + \mathbf{d}'\mathbf{z}(\mathbf{x})\\ &\quad + 2d_0 f_0(\mathbf{x})]\,d\mathbf{x}.\end{aligned}$$

This inequality suggests that

$$f_0(\mathbf{x}) = m(\mathbf{x})\mathbf{z}'(\mathbf{x})\mathbf{H}^{-1}(\xi)\boldsymbol{\beta} + \mathbf{z}'(\mathbf{x})\mathbf{c}$$

for some p-vectors $\beta$ [proportional to $b(f_0,\xi)$] and $\mathbf{c}$. Choosing $\mathbf{c}$ to satisfy $\int_S \mathbf{z}'(\mathbf{x})f(\mathbf{x}) = \mathbf{0}$, we have

$$f_0(\mathbf{x}) = \mathbf{z}'(\mathbf{x})[m(\mathbf{x})\mathbf{H}^{-1}(\xi) - B^{-1}]\boldsymbol{\beta}. \tag{2.6}$$

To see that there in fact exists a bias-maximizing $f_0$, and that it has the form (2.6) for some $\beta$, define

$$h(\mathbf{x}, \boldsymbol{\beta}) = s\mathbf{z}'(\mathbf{x})[m(\mathbf{x})\mathbf{H}^{-1}(\xi) - B^{-1}]\boldsymbol{\beta},$$

where $s > 0$ is chosen so that

$$\int_S h(\mathbf{x}, \boldsymbol{\beta})\, d\mathbf{x} = \eta^2. \tag{2.7}$$

Thus $h(\mathbf{x}, \boldsymbol{\beta}) \in \mathcal{F}_1$ for any $\beta$. For a fixed but arbitrary $f \in \mathcal{F}_1$, consider $h(\mathbf{x}, \mathbf{b})$, with $\mathbf{b}$ as in (2.4). We assert that

$$B(f, \xi) \leq B(h(\mathbf{x}, \mathbf{b}), \xi), \tag{2.8}$$

so that we may restrict the search for a maximizing $f_0$ to those functions of the form (2.6).

Furthermore, we will show that equality holds in (2.8) if and only if

$$f(\mathbf{x}) = h(\mathbf{x}, \mathbf{b}) \qquad a.e \ \ \mathbf{x}. \tag{2.9}$$

Using (2.7), we have

$$s^2 = \frac{\eta^2}{\mathbf{b}'[\mathbf{H}^{-1}(\xi)\mathbf{K}(\xi) - \mathbf{H}^{-1}(\xi)]\mathbf{b}}. \tag{2.10}$$

Replacing $f$ by $h(\mathbf{x}, \mathbf{b})$ in $\mathbf{b}(f, \xi)$, we have

$$\mathbf{b}(h(\mathbf{x}, \mathbf{b}), \xi) = s[\mathbf{K}(\xi)\mathbf{H}^{-1}(\xi) - \mathbf{I}]\mathbf{b}(f, \xi)$$

and

$$B(h(\mathbf{x}, \mathbf{b}(f, \xi)), \xi) = s^2\mathbf{b}'(f, \xi)[\mathbf{H}^{-1}(\xi)\mathbf{K}(\xi) - \mathbf{I}]\mathbf{H}^{-1}(\xi)[\mathbf{K}(\xi)\mathbf{H}^{-1}(\xi) - \mathbf{I}]\mathbf{b}(f, \xi).$$

Since $f \in \mathcal{F}_1$,

$$\begin{aligned} s^2[\mathbf{b}'(f,\xi)\mathbf{H}^{-1}(\boldsymbol{\xi})\mathbf{b}(f,\xi)]^2 &= \left(\int_S f(\mathbf{x})h(\mathbf{x},\mathbf{b}(f,\xi))\,d\mathbf{x}\right)^2 \\ &\leq \int_S f^2(\mathbf{x})\,d\mathbf{x}\int_S h^2(\mathbf{x},\mathbf{b}(f,\xi))\,d\mathbf{x} \\ &= \eta^4. \end{aligned} \tag{2.11}$$

Let $H^{-1/2}(\xi)$ be a symmetric, positive root of $H^{-1}(\xi)$, and let

$$\begin{aligned} \mathbf{l} &= H^{-1/2}(\xi)\mathbf{b}(f,\xi), \\ \mathbf{J} &= H^{-1/2}(\xi)K(\xi)H^{-1/2}(\xi) - \mathbf{I}. \end{aligned}$$

Represent $\mathbf{J}$ as $\mathbf{J} = \mathbf{Q}\Lambda\mathbf{Q}'$, where $\mathbf{Q}$ is orthogonal and $\Lambda$ is diagonal. Define

$$\mathbf{u} = \frac{\mathbf{Q}'\mathbf{l}}{(\mathbf{l}'\mathbf{l})^{1/2}}.$$

Note that $\mathbf{u}'\mathbf{u} = 1$, and put $\bar{\lambda} = \sum_{i=1}^N \lambda_i u_i^2$, where $\lambda_i = \Lambda_{ii}$. In this notation, we have

$$B(h(\mathbf{x},\mathbf{b}(f,\xi)),\xi) = s^2\mathbf{l}'\mathbf{J}^2\mathbf{l}, \tag{2.12}$$

$$B(f,\xi) = \mathbf{l}'\mathbf{l}. \tag{2.13}$$

From (2.10) and (2.11),

$$\begin{aligned} s^2 &= \frac{\eta^2}{\mathbf{l}'\mathbf{J}\mathbf{l}}, \text{ and} \\ \eta^2 &\geq s\mathbf{l}'\mathbf{l}, \end{aligned}$$

so that

$$\eta^2 \geq \frac{(\mathbf{l}'\mathbf{l})^2}{\mathbf{l}'\mathbf{J}\mathbf{l}} = \frac{\mathbf{l}'\mathbf{l}}{\mathbf{u}'\Lambda\mathbf{u}}. \tag{2.14}$$

Now (2.12) - (2.14) give

$$\begin{aligned}\frac{B(h(\mathbf{x},\mathbf{b}(f,\xi)),\xi)}{B(f,\xi)} - 1 &= \frac{\eta^2}{\mathbf{l}'\mathbf{l}} \cdot \frac{\mathbf{u}'\Lambda^2\mathbf{u}}{\mathbf{u}'\Lambda\mathbf{u}} - 1 \\ &\geq \frac{\mathbf{u}'\Lambda^2\mathbf{u}}{(\mathbf{u}'\Lambda\mathbf{u})^2} - 1 \\ &= \frac{\sum_{i=1}^{N}(\lambda_i - \bar{\lambda})u_i^2}{\bar{\lambda}^2} \\ &\geq 0.\end{aligned}$$

This proves (2.8). Equality in (2.14) requires equality in (2.11), and hence that (2.9) implies equality in (2.14). Thus, equality in (2.8) is equivalent to (2.9 ). Rather than investigate solutions to (2.9), it is now simpler to determine an $f_0$ of the form (2.6), with $\boldsymbol{\beta}$ chosen to maximize

$$B(f_0,\xi) = \boldsymbol{\beta}'(\mathbf{K}(\xi)\mathbf{H}^{-1}(\xi) - \mathbf{I})'\mathbf{H}^{-1}(\xi)(\mathbf{K}(\xi)\mathbf{H}^{-1}(\xi) - \mathbf{I})\boldsymbol{\beta},$$

subject to

$$\int_S f_0^2(\mathbf{x})\,d\mathbf{x} = \boldsymbol{\beta}'(\mathbf{H}^{-1}(\xi)\mathbf{K}(\xi)\mathbf{H}^{-1}(\xi) - \mathbf{H}^{-1}(\xi))\boldsymbol{\beta} = \eta^2.$$

Equivalently, with

$$\begin{aligned}\boldsymbol{\delta} &= \mathbf{K}^{-1}(\xi)\boldsymbol{\beta}, \\ \mathbf{G} &= (\mathbf{K}(\xi)\mathbf{H}^{-1}(\xi) - \mathbf{I})\mathbf{K}(\xi)\mathbf{H}^{-1}\mathbf{K}(\xi), \\ \mathbf{F} &= \mathbf{G}(\mathbf{H}(\xi)\mathbf{K}^{-1}(\xi) - \mathbf{I}),\end{aligned}$$

we maximize

$$B(f_0,\xi) = \boldsymbol{\delta}'\mathbf{F}\boldsymbol{\delta}$$

subject to

$$\boldsymbol{\delta}'\mathbf{G}\boldsymbol{\delta} = \eta^2.$$

This is a standard eigenvalue problem. The solution is obtained by finding the largest solution $\mu(\xi)$ to $|\mathbf{F} - \mu\mathbf{G}| = 0$, and then choosing $\boldsymbol{\delta}$ to satisfy

$$(\mathbf{F} - \mu(\xi)\mathbf{G})\boldsymbol{\delta} = 0,$$

normalized to satisfy $\boldsymbol{\delta}'\mathbf{G}\boldsymbol{\delta} = \eta^2$. In terms of $\mathbf{H}(\xi)$and $\mathbf{K}(\xi)$, $v(\xi) = 1 + \mu(\xi)$ is the largest solution to

$$|\mathbf{K}(\xi) - v(\xi)\mathbf{H}(\xi)| = 0,$$

and

$$(\mathbf{K}(\xi) - v(\xi)\mathbf{H}(\xi))\boldsymbol{\delta} = 0.$$

Hence,

$$max_{\mathcal{F}_1} B(f, \xi) = B(f_0, \xi) = \mu(\xi)\eta^2,$$

and thus,

$$\max_{\mathcal{F}} \mathcal{L}_Q(f, \xi) = \eta^2[\nu tr\{\mathbf{B}^{-1}(\xi) \cdot \mathbf{A}\} + ch_{max}(\mathbf{K}(\xi)\mathbf{H}^{-1}(\xi))]. \qquad (2.15)$$

Thus the density $m(\mathbf{x})$ of a Q-optimal (minimax) deign $\xi$ must minimize the right hand side of (2.15).

## 2.3 Other Overall Measures of Loss

In addition to the Integrated Mean Squared Error, we shall be using the Determinant of the Mean Squared Error Matrix and Trace of the Mean Squared Error Matrix as measures of loss. We shall construct optimal designs for the two interacting regressors model under each of these measures of overall loss.

In equation (2.9) the Mean Squared Error Matrix is given by

$$MSE(f,\xi) = \frac{\sigma_\varepsilon^2}{n}\mathbf{B}^{-1}(\xi) + \mathbf{B}^{-1}(\xi)b(f,\xi)\mathbf{b}'(f,\xi)\mathbf{B}^{-1}(\xi).$$

Thus the determinant of the Mean Squared Error Matrix is given by

$$\begin{aligned}
\mathcal{L}_D(f,\xi) &= det\{MSE(f,\xi)\} \\
&= det\{\frac{\sigma_\varepsilon^2}{n}\mathbf{B}^{-1}(\xi) + \mathbf{B}^{-1}(\xi)\mathbf{b}(f,\xi)\mathbf{b}'(f,\xi)\mathbf{B}^{-1}(\xi)\} \\
&= \left(\frac{\sigma_\varepsilon^2}{n}\right)^p det\{\mathbf{B}^{-1}(\xi)[\mathbf{I} + \frac{n}{\sigma_\varepsilon^2}\mathbf{b}(f,\xi)\mathbf{b}'(f,\xi)\mathbf{B}^{-1}(\xi)]\} \\
&= \left(\frac{\sigma_\varepsilon^2}{n}\right)^p det\{\mathbf{B}^{-1}(\xi)\}det\{\mathbf{I} + \frac{n}{\sigma_\varepsilon^2}\mathbf{b}(f,\xi)\mathbf{b}'(f,\xi)\mathbf{B}^{-1}(\xi)\} \\
&= \left(\frac{\sigma_\varepsilon^2}{n}\right)^p \frac{1}{|\mathbf{B}(\xi)|}\left\{1 + \frac{n}{\sigma_\varepsilon^2}\mathbf{b}'(f,\xi)\mathbf{B}^{-1}(\xi)\mathbf{b}(f,\xi)\right\}
\end{aligned}$$

The trace of the Mean Squared Error Matrix is given by

$$\begin{aligned}
\mathcal{L}_A(f,\xi) &= tr\{MSE(f,\xi)\} \\
&= \frac{\sigma_\varepsilon^2}{n}tr\{\mathbf{B}^{-1}(\xi)\} + tr\{\mathbf{B}^{-1}(\xi)\mathbf{b}(f,\xi)\mathbf{b}'(f,\xi)\mathbf{B}^{-1}(\xi)\} \\
&= \frac{\sigma_\varepsilon^2}{n}tr\{\mathbf{B}^{-1}(\xi)\} + tr\{\mathbf{b}'(f,\xi)\mathbf{B}^{-2}(\xi)\mathbf{b}(f,\xi)\} \\
&= \frac{\sigma_\varepsilon^2}{n}tr\{\mathbf{B}^{-1}(\xi)\} + \mathbf{b}'(f,\xi)\mathbf{B}^{-2}(\xi)\mathbf{b}(f,\xi).
\end{aligned}$$

The maxima of the determinant and trace of the Mean Squared Error Matrix can be derived as was that of the Integrated Mean Squared Error, and are given by

$$\begin{aligned}
max_{\mathcal{F}}\mathcal{L}_D(f,\xi) &= \eta^2\left(\frac{\sigma_\varepsilon^2}{n}\right)^{p-1}\frac{1}{|\mathbf{B}(\xi)|}\left(\nu + ch_{max}(\mathbf{G}(\xi)\mathbf{B}^{-1}(\xi))\right), \\
max_{\mathcal{F}}\mathcal{L}_A(f,\xi) &= \eta^2\left(\nu tr(\mathbf{B}^{-1}(\xi)) + ch_{max}\{\mathbf{G}(\xi)\mathbf{B}^{-2}(\xi)\}\right),
\end{aligned}$$

respectively. See Wiens (1992) for details.

## 2.4 Minimizing Maximum Loss over the Designs $\xi$

The "minimax" approach to robust optimal designs is a two-phase problem. In the last section we have maximized the loss over $\mathcal{F}$ using the Integrated Mean Squared Error, Determinant of the Mean Squared Error Matrix and Trace of the Mean Square Error Matrix. In each case, we have derived the maximum loss function. The phase we are at now is to minimize these maximum loss functions over the designs $\xi$. In tackling this problem it appears necessary to consider individual models.

As stated earlier, we are interested in the model

$$Y = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_{12} x_1 x_2 + f(\mathbf{x}) + \varepsilon \tag{2.16}$$

on

$$\begin{aligned} \mathcal{S} &= [-1/2, 1/2] \times [-1/2, 1/2] \\ \mathbf{x} &= (x_1, x_2)' \\ z(\mathbf{x}) &= (1, x_1, x_2, x_1 x_2)' \end{aligned}$$

with two interacting regressors.

The model, by its definition, is exchangeable in $x_1$ and $x_2$ (with appropriate reparameterization of the unknown $\theta = (\theta_0, \theta_1, \theta_2, \theta_{12})'$. It makes sense to assume that the optimal density $m(\mathbf{x})$ is symmetric and exchangeable in $x_1$ and $x_2$. Thus all integrals of the form $\int_{\mathcal{S}} x_i x_j m(\mathbf{x}) d\mathbf{x}$ $(i \neq j)$ vanish.

For this model we set out to construct the loss functions corresponding to each of

$$max_{\mathcal{F}}\mathcal{L}_Q(f,\xi) = \eta^2[\nu tr\{\mathbf{B}^{-1}(\xi)\cdot\mathbf{A}\} + ch_{max}\{\mathbf{K}(\xi)\mathbf{H}^{-1}(\xi)\}], \quad (2.17)$$

$$max_{\mathcal{F}}\mathcal{L}_D(f,\xi) = \eta^2\left(\frac{\sigma_\varepsilon^2}{n}\right)^3 \frac{[\nu + ch_{max}(\mathbf{G}(\boldsymbol{\xi})\mathbf{B}^{-1}(\xi))]}{|\mathbf{B}(\xi)|}, \text{ and} \quad (2.18)$$

$$max_{\mathcal{F}}\mathcal{L}_A(f,\xi) = \eta^2[\nu tr(\mathbf{B}^{-1}(\xi)) + ch_{max}\{\mathbf{G}(\boldsymbol{\xi})\mathbf{B}^{-2}(\xi)\}]. \quad (2.19)$$

We find

$$\begin{aligned}
\mathbf{A} &= \int_S \mathbf{z}(\mathbf{x})\mathbf{z}'(\mathbf{x})\,d\mathbf{x} \\
&= \int_{-\frac{1}{2}}^{\frac{1}{2}}\int_{-\frac{1}{2}}^{\frac{1}{2}} \begin{pmatrix} 1 & x_1 & x_2 & x_1x_2 \\ x_1 & x_1^2 & x_1x_2 & x_1^2x_2 \\ x_2 & x_1x_2 & x_2^2 & x_1x_2^2 \\ x_1x_2 & x_1^2x_2 & x_1x_2^2 & x_1^2x_2^2 \end{pmatrix} dx_1dx_2 \\
&= diag\left(1, \frac{1}{12}, \frac{1}{12}, \frac{1}{144}\right), \\
\mathbf{B} &= \int_S \mathbf{z}(\mathbf{x})\mathbf{z}'(\mathbf{x})m(\mathbf{x})\,d\mathbf{x} \\
&= diag\left(1, \gamma, \gamma, \gamma_{12}\right),
\end{aligned}$$

where

$$\begin{aligned}
\gamma &= \int_S x_1^2 m(\mathbf{x})\,d\mathbf{x}, \\
\gamma_{12} &= \int_S x_1^2x_2^2 m(\mathbf{x})\,d\mathbf{x}.
\end{aligned}$$

$$
\begin{aligned}
\mathbf{K}(\xi) &= \int_S \mathbf{z}(\mathbf{x})\mathbf{z}'(\mathbf{x})m^2(\mathbf{x})\,d\mathbf{x} \\
&= diag(k_1, k_2, k_3, k_4) \text{ where} \\
k_1 &= \int_S m^2(\mathbf{x})\,d\mathbf{x}, \\
k_2 &= \int_S x_1^2 m^2(\mathbf{x})\,d\mathbf{x}, \\
k_3 &= \int_S x_2^2 m^2(\mathbf{x})\,d\mathbf{x}, \\
k_4 &= \int_S x_1^2 x_2^2 m^2(\mathbf{x})\,d\mathbf{x}.
\end{aligned}
$$

$$
\begin{aligned}
\mathbf{H}(\xi) &= \mathbf{B}(\xi)\mathbf{A}^{-1}\mathbf{B}(\xi) \\
&= diag\left(1, 12\gamma^2, 12\gamma^2, 144\gamma_{12}^2\right)
\end{aligned}
$$

$$
\begin{aligned}
\mathbf{G}(\xi) &= \mathbf{K}(\xi) - \mathbf{H}(\xi) \\
&= diag\left(k_1 - 1, k_2 - 12\gamma^2,\right. \\
&\quad \left. k_3 - 12\gamma^2, k_4 - 144\gamma_{12}^2\right)
\end{aligned}
$$

$$\begin{aligned}
\mathbf{K}(\xi)\mathbf{H}^{-1}(\xi) &= \left(\int_S \mathbf{z}(\mathbf{x})\mathbf{z}'(\mathbf{x})m^2(\mathbf{x})\,d\mathbf{x}\right)\mathbf{H}^{-1}(\xi) \\
&= diag\left(k_1, \frac{k_2}{12\gamma^2}, \frac{k_3}{12\gamma^2}, \frac{k_4}{144\gamma_{12}^2}\right)
\end{aligned}$$

$$\begin{aligned}
\mathbf{G}(\xi)\mathbf{B}^{-1}(\xi) &= (\mathbf{K}(\xi) - \mathbf{H}(\xi))\mathbf{B}^{-1}(\xi) \\
&= diag\left(k_1 - 1, \frac{k_2}{\gamma} - 12\gamma,\right. \\
&\qquad \left.\frac{k_3}{\gamma} - 12\gamma, \frac{k_4}{\gamma_{12}} - 144\gamma_{12}\right)
\end{aligned}$$

$$\begin{aligned}
\mathbf{G}(\xi)\mathbf{B}^{-2}(\xi) &= (\mathbf{K}(\xi) - \mathbf{H}(\xi))\mathbf{B}^{-2}(\xi) \\
&= diag\left(k_1 - 1, \frac{k_2}{\gamma^2} - 12,\right. \\
&\qquad \left.\frac{k_3}{\gamma^2} - 12, \frac{k_4}{\gamma_{12}^2} - 144\right)
\end{aligned}$$

Using (2.17)-(2.19), we have

$$\begin{aligned}
max_{\mathcal{F}}\mathcal{L}_Q(f,\xi) &= \eta^2\left\{\nu(1 + \frac{1}{6\gamma} + \frac{1}{144\gamma_{12}}) + v_Q(\xi)\right\} \\
max_{\mathcal{F}}\mathcal{L}_D(f,\xi) &= \eta^2\left(\frac{\sigma_\varepsilon^2}{n}\right)^3 \frac{(\nu + v_D(\xi))}{\gamma^2\gamma_{12}} \\
max_{\mathcal{F}}\mathcal{L}_A(f,\xi) &= \eta^2\left\{\nu(1 + \frac{2}{\gamma} + \frac{1}{\gamma_{12}}) + v_A(\xi)\right\}
\end{aligned}$$

where

$$\begin{aligned} v_Q(\xi) &= ch_{max}(\mathbf{K}(\xi)\mathbf{H}^{-1}(\xi)) \\ &= \max\left(\int_S m^2(\mathbf{x})\,d\mathbf{x}, \frac{1}{12\gamma^2}\int_S {x_1}^2 m^2(\mathbf{x})\,d\mathbf{x}, \right. \\ &\quad \left. \frac{1}{144\gamma_{12}^2}\int_S x_1^2 x_2^2 m^2(\mathbf{x})\,d\mathbf{x}\right) \end{aligned}$$

$$\begin{aligned} v_D(\xi) &= ch_{max}(\mathbf{G}(\xi)\mathbf{B}^{-1}(\xi)) \\ &= \max\left(\int_S m^2(\mathbf{x})\,d\mathbf{x}-1, \frac{1}{\gamma}\int_S x_1^2 m^2(\mathbf{x})\,d\mathbf{x}-12\gamma, \right. \\ &\quad \left. \frac{1}{\gamma_{12}}\int_S x_1^2 x_2^2 m^2(\mathbf{x})\,d\mathbf{x} - 144\gamma_{12}\right) \end{aligned}$$

$$\begin{aligned} v_A(\xi) &= ch_{max}(\mathbf{G}(\xi)\mathbf{B}^{-2}(\xi)) \\ &= \max\left(\int_S m^2(\mathbf{x})\,d\mathbf{x}-1, \frac{1}{\gamma^2}\int_S x_1^2 m^2(\mathbf{x})\,d\mathbf{x}-12, \right. \\ &\quad \left. \frac{1}{\gamma_{12}^2}\int_S x_1^2 x_2^2 m(\mathbf{x})^2\,d\mathbf{x} - 144\right) \end{aligned}$$

# Chapter 3

# ANALYTICAL AND NUMERICAL RESULTS

In this chapter we present optimal densities using the Integrated Mean Squared Error, the Determinant of the Mean Squared Error and Trace of the Mean Squared Error as the overall measure of loss. Wiens (1990) reported the optimal density corresponding to small values of $\nu$ for Integrated Mean Squared Error for model (2.16). Here, in addition to results for small values of $\nu$ for the Integrated Mean Squared Error, we present results for larger values of $\nu$. In fact, we would assert that the optimal density for large values of $\nu$ retains the same form as for small values, but the coefficient are no longer nonnegative. Furthermore, we shall present optimal densities corresponding to Determinant Mean Squared Error loss and Trace of Mean Squared Error loss.

Following the approach of Heo, Schmuland and Wiens (2001), our design will depend on $\eta^2$ and $\sigma^2$ only through $\nu = \sigma^2/(n\eta^2)$, which may be chosen by the experimenter according to her judgement of the relative importance of variance versus bias. Heo, Schmuland and Wiens (2001) gave an alternative explanation, that this parameter is inversely related to the premium, in

terms of lost efficiency relative to the variance-minimizing design, that the experimenter is willing to pay for robustness against model misspecification.

We observed that the form of the optimal density is determined by

$$\begin{aligned} v_Q(\xi) &= ch_{max}(\mathbf{K}(\xi)\mathbf{H}^{-1}(\xi)) \\ v_D(\xi) &= ch_{max}(\mathbf{G}(\xi)\mathbf{B}^{-1}(\xi)) \\ v_Q(\xi) &= ch_{max}(\mathbf{G}(\xi)\mathbf{B}^{-2}(\xi)) \end{aligned}$$

for Integrated Mean Squared Error loss, Determinant of the Mean Squared Error loss and Trace of the Mean Squared Error loss respectively. Each of $\mathbf{K}(\xi)\mathbf{H}^{-1}(\xi), \mathbf{G}(\xi)\mathbf{B}^{-1}(\xi)$ and $\mathbf{G}(\xi)\mathbf{B}^{-2}(\xi)$ are diagonal matrices with three distinct entries. We shall present the corresponding optimal density when each of the three distinct eigenvalues assumes maximum.

## 3.1 Optimal Density Corresponding to First Eigenvalues

Suppose for each of $\mathbf{K}(\xi)\mathbf{H}^{-1}(\xi), \mathbf{G}(\xi)\mathbf{B}^{-1}(\xi)$ and $\mathbf{G}(\xi)\mathbf{B}^{-2}(\xi)$ maximum eigenvalue is the first, that is

$$\begin{aligned} v_Q(\xi) &= ch_{max}(\mathbf{K}(\xi)\mathbf{H}^{-1}(\xi)) = \int_S m^2(\mathbf{x})d\mathbf{x} \\ v_D(\xi) &= ch_{max}(\mathbf{G}(\xi)\mathbf{B}^{-1}(\xi)) = \int_S m^2(\mathbf{x})d\mathbf{x} - 1 \\ v_Q(\xi) &= ch_{max}(\mathbf{G}(\xi)\mathbf{B}^{-2}(\xi)) = \int_S m^2(\mathbf{x})d\mathbf{x} - 1. \end{aligned}$$

We find density $m(\mathbf{x})$ minimizing $max_{\mathcal{F}}\mathcal{L}_Q(f,\xi)$, $max_{\mathcal{F}}\mathcal{L}_D(f,\xi)$ and $max_{\mathcal{F}}\mathcal{L}_A(f,\xi)$.

Let

$$\begin{aligned} L_Q(\xi) &= \eta^{-2} max_{\mathcal{F}}\mathcal{L}_Q(f,\xi), \\ L_D(\xi) &= \eta^{-2}(\frac{\sigma_\varepsilon^2}{n})^{-3} max_{\mathcal{F}}\mathcal{L}_D(f,\xi), \\ L_A(\xi) &= \eta^{-2} max_{\mathcal{F}}\mathcal{L}_A(f,\xi). \end{aligned}$$

Let $\gamma$ and $\gamma_{12}$ be fixed.

For fixed $\gamma, \gamma_{12}$ we minimize

$$\begin{aligned} L_Q(\xi) &= \nu\left(1+\frac{1}{6\gamma}+\frac{1}{144\gamma_{12}}\right)+\int_S m^2(\mathbf{x})d\mathbf{x}, \\ L_D(\xi) &= \frac{(\nu+\int_S m^2(\mathbf{x})d\mathbf{x}-1)}{\gamma^2\gamma_{12}}, \\ L_A(\xi) &= \nu\left(1+\frac{2}{\gamma}+\frac{1}{\gamma_{12}}\right)+\int_S m^2(\mathbf{x})d\mathbf{x}-1, \end{aligned}$$

subject to

$$\int_S m(\mathbf{x})d\mathbf{x} = 1, \tag{3.1}$$

$$\int_S x_1^2 m(\mathbf{x})\,d\mathbf{x} = \gamma, \tag{3.2}$$

$$\int_S x_2^2 m(\mathbf{x})\,d\mathbf{x} = \gamma, \tag{3.3}$$

$$\int_S x_1^2 x_2^2 m(\mathbf{x})\,d\mathbf{x} = \gamma_{12}. \tag{3.4}$$

Equivalently, for the three cases we minimize

$$\int_S m^2(\mathbf{x})d\mathbf{x}$$

subject to (3.1) - (3.4).

It is sufficient if $m_0$ minimizes

$$\int_S m^2(\mathbf{x})\,d\mathbf{x} - 2c\int_S x_1^2x_2^2m(\mathbf{x})\,d\mathbf{x} - 2b\int_S (x_1^2+x_2^2)m(\mathbf{x})\,d\mathbf{x}$$
$$-2a\int_S m(\mathbf{x})\,d\mathbf{x}$$

for some constants $a, b$ and $c$ and satisfies the side conditions. This is since in the class of $m's$ satisfying these conditions,

$$-2c\int_S x_1^2x_2^2m(\mathbf{x})\,d\mathbf{x} - 2b\int_S (x_1^2+x_2^2)m(\mathbf{x})\,d\mathbf{x} - 2a\int_S m(\mathbf{x})\,d\mathbf{x} = \textit{constant}.$$

Thus such an $m_0$ will solve the original constrained problem.

Equivalently, we minimize

$$\int_S \left[m^2(\mathbf{x}) - 2c\,x_1^2x_2^2m(\mathbf{x}) - 2b(x_1^2+x_2^2)m(\mathbf{x}) - 2am(\mathbf{x})\right]d\mathbf{x}$$

by minimizing the integrand

$$[y^2 - 2(a + b(x_1^2+x_2^2) + cx_1^2x_2^2)y]|_{y=m(\mathbf{x})},$$

pointwise.

Now $y^2 - 2(a + b(x_1^2+x_2^2) + cx_1^2x_2^2)y$ is minimized over $\{y \geq 0\}$ at

$$(a + b(x_1^2+x_2^2) + cx_1^2x_2^2)^+$$

where

$$k^+ = \begin{cases} k & \text{if } k \geq 0 \\ 0 & \text{if } k < 0 \end{cases}$$

Thus,

$$m_0(\mathbf{x}) = (a + b(x_1^2+x_2^2) + cx_1^2x_2^2)^+ \tag{3.5}$$

with $a, b, c$ determined by (3.1) - (3.4).

## 3.2 Optimal Density Corresponding to Second Eigenvalues

Suppose for the three cases the maximum eigenvalue is the second, that is

$$\begin{aligned} v_Q(\xi) &= ch_{max}(\mathbf{K}(\xi)\mathbf{H}^{-1}(\xi)) = \frac{1}{12\gamma^2}\int_S x_1^2m^2(\mathbf{x})d\mathbf{x} \\ v_D(\xi) &= ch_{max}(\mathbf{G}(\xi)\mathbf{B}^{-1}(\xi)) = \frac{1}{\gamma}\int_S x_1^2m^2(\mathbf{x})d\mathbf{x} - 12\gamma \\ v_A(\xi) &= ch_{max}(\mathbf{G}(\xi)\mathbf{B}^{-2}(\xi)) = \frac{1}{\gamma^2}\int_S x_1^2m^2(\mathbf{x})d\mathbf{x} - 12 \end{aligned}$$

Now, we find the density $m(\mathbf{x})$ minimizing

$$\begin{aligned} L_Q(\xi) &= \nu\left(1+\frac{1}{6\gamma}+\frac{1}{144\gamma_{12}}\right)+\frac{1}{12\gamma^2}\int_S x_1^2 m^2(\mathbf{x})d\mathbf{x} \\ L_D(\xi) &= \frac{(\nu+\frac{1}{\gamma}\int_S x_1^2 m^2(\mathbf{x})d\mathbf{x}-12\gamma)}{\gamma^2\gamma_{12}} \\ L_A(\xi) &= \nu\left(1+\frac{2}{\gamma}+\frac{1}{\gamma_{12}}\right)+\frac{1}{\gamma^2}\int_S x_1^2 m^2(\mathbf{x})d\mathbf{x}-12 \end{aligned}$$

subject to (3.1) - (3.4). Equivalently, we minimize

$$\int_S (x_1^2+x_2^2)m^2(\mathbf{x})d\mathbf{x}$$

subject to (3.1) - (3.4).

It is sufficient if $m_0$ minimizes

$$\int_S (x_1^2+x_2^2)m^2(\mathbf{x})\,d\mathbf{x}-2c\int_S x_1^2x_2^2 m(\mathbf{x})\,d\mathbf{x}-2b\int_S (x_1^2+x_2^2)m(\mathbf{x})\,d\mathbf{x}$$
$$-2a\int_S m(\mathbf{x})\,d\mathbf{x}$$

for some constants $a, b$ and $c$ and satisfies the side conditions. In the class of $m's$ satisfying these conditions,

$$-2c\int_S x_1^2x_2^2 m(\mathbf{x})\,d\mathbf{x}-2b\int_S (x_1^2+x_2^2)m(\mathbf{x})\,d\mathbf{x}-2a\int_S m(\mathbf{x})\,d\mathbf{x}=\textit{constant}.$$

Thus such an $m_0$ will solve the original constrained problem. Equivalently, we minimize

$$\int_S \left[(x_1^2+x_2^2)m^2(\mathbf{x})-2c\,x_1^2x_2^2m(\mathbf{x})-2b(x_1^2+x_2^2)m(\mathbf{x})\;-2am(\mathbf{x})\right]d\mathbf{x}$$

by minimizing the integrand

$$\left[(x_1^2+x_2^2)y^2-2(a+b(x_1^2+x_2^2)+cx_1^2x_2^2)y\right]\Big|_{y=m(\mathbf{x})}$$

pointwise.

Now $(x_1^2+x_2^2)y^2-2(a+b(x_1^2+x_2^2)+cx_1^2x_2^2)y$ is minimized over $\{y\geq 0\}$ at

$$\frac{(a+b(x_1^2+x_2^2)+cx_1^2x_2^2)^+}{x_1^2+x_2^2}$$

Thus,

$$m_0(\mathbf{x})=\frac{(a+b(x_1^2+x_2^2)+cx_1^2x_2^2)^+}{x_1^2+x_2^2}$$

with $a,b,c$ determined by (3.1) - (3.4).

## 3.3 Optimal Density Corresponding to Third Eigenvalues

Suppose for the three cases the maximum eigenvalue is the third, that is

$$\begin{aligned}
\upsilon_Q(\xi) &= ch_{max}(\mathbf{K}(\xi)\mathbf{H}^{-1}(\xi))=\frac{1}{144\gamma_{12}^2}\int_S x_1^2x_2^2m^2(\mathbf{x})d\mathbf{x}\\
\upsilon_D(\xi) &= ch_{max}(\mathbf{G}(\xi)\mathbf{B}^{-1}(\xi))=\frac{1}{\gamma_{12}}\int_S x_1^2x_2^2m^2(\mathbf{x})d\mathbf{x}-144\gamma\\
\upsilon_A(\xi) &= ch_{max}(\mathbf{G}(\xi)\mathbf{B}^{-2}(\xi))=\frac{1}{\gamma_{12}^2}\int_S x_1^2x_2^2m^2(\mathbf{x})d\mathbf{x}-144
\end{aligned}$$

Working through the same process as in the last two section we obtained

$$m_0(\mathbf{x})=\frac{(a+b(x_1^2+x_2^2)+cx_1^2x_2^2)^+}{x_1^2x_2^2} \tag{3.6}$$

with $a, b, c$ determined by (3.1) - (3.4).

These results agree with the work of Shi (2002). He reported that the minimax density corresponding to $max_{\mathcal{F}}\mathcal{L}_Q(f,\xi)$, $max_{\mathcal{F}}\mathcal{L}_D(f,\xi)$ and $max_{\mathcal{F}}\mathcal{L}_A(f,\xi)$ has the form

$$m_0(\mathbf{x}) = \left( \frac{\sum_{i,j=1}^{p} a_{ij}\mathbf{z_i}(\mathbf{x})\mathbf{z_j}(\mathbf{x}) + a_{00}}{\sum_{i,j=1}^{p} b_{ij}\mathbf{z_i}(\mathbf{x})\mathbf{z_j}(\mathbf{x})} \right)^{+}, \mathbf{x} \in S$$

where the coefficients $a_{00}, a_{ij}, b_{ij}$, for $i, j = 1, 2, ..., p$ are some constants. In particular, for the regression model (2.16), which we are attempting to proffer solutions

$$m_0(\mathbf{x}) = \left( \frac{a + b(x_1^2 + x_2^2) + cx_1^2x_2^2}{d + e(x_1^2 + x_2^2) + fx_1^2x_2^2} \right)^{+}. \tag{3.7}$$

The form (3.7) has the advantage that it does not depend on the maximum eigenvalue. However, as we shall see later, it poses some numerical difficulties due to its over-parametrization.

## 3.4 Optimal Coefficients for small $\nu$

Having derived the forms of the optimal densities corresponding to the various choices of overall measure of loss in the previous sections, we set out to determine the coefficients for optimum designs. We shall proceed analytically as far as we can and resort to numerical methods where we do not have closed form solutions.

Suppose

$$\begin{aligned} ch_{max}(\mathbf{K}(\xi)\mathbf{H}^{-1}(\xi)) &= \int_S m^2(\mathbf{x})d\mathbf{x}, \\ ch_{max}(\mathbf{G}(\xi)\mathbf{B}^{-1}(\xi)) &= \int_S m^2(\mathbf{x})d\mathbf{x} - 1, \\ ch_{max}(\mathbf{G}(\xi)\mathbf{B}^{-2}(\xi)) &= \int_S m^2(\mathbf{x})d\mathbf{x} - 1. \end{aligned}$$

From the results of the previous sections, the optimal density is of the form

$$m_0(\mathbf{x}) = (a + b(x_1^2 + x_2^2) + cx_1^2x_2^2)^+$$

with $a, b, c$ determined by

$$\int_S m_0(\mathbf{x})d\mathbf{x} = 1, \int_S x_1^2 m_0(\mathbf{x})\, d\mathbf{x} = \gamma, \int_S x_1^2x_2^2 m_0(\mathbf{x})\, d\mathbf{x} = \gamma_{12}. \tag{3.8}$$

If $a, b$ and $c$ are nonnegative, then the optimal density simplifies to

$$m_0(\mathbf{x}) = (a + b(x_1^2 + x_2^2) + cx_1^2x_2^2). \tag{3.9}$$

Plugging equation (3.9) into (3.8), we have

$$\begin{aligned} a &= \frac{81}{16} - \frac{135}{2}\gamma + 225\gamma_{12}, && (3.10)\\ b &= \frac{-135}{4} + 630\gamma - 2700\gamma_{12}, && (3.11)\\ c &= 225 - 5400\gamma + 32400\gamma_{12}. && (3.12) \end{aligned}$$

In order to choose $\gamma$ and $\gamma_{12}$ such that $a, b, c \geq 0$, the following inequalities should be satisfied:

$$\begin{aligned} \gamma_{12} &\geq \frac{3}{10}\gamma - \frac{9}{400}, \\ \gamma_{12} &\leq \frac{7}{30}\gamma - \frac{1}{80}, \\ \gamma_{12} &\geq \frac{1}{6}\gamma - \frac{1}{144}. \end{aligned}$$

This implies

$$\gamma \leq \frac{7}{60}, \frac{1}{6}\gamma - \frac{1}{144} \leq \gamma_{12} \leq \frac{7}{30}\gamma - \frac{1}{80},$$

or

$$\gamma \geq \frac{7}{60}, \frac{3}{10}\gamma - \frac{9}{400} \leq \gamma_{12} \leq \frac{7}{30}\gamma - \frac{1}{80}.$$

Now we set to minimize

$$\begin{aligned} L_Q(\xi) &= \nu(1 + \frac{1}{6\gamma} + \frac{1}{144\gamma_{12}}) + \int_S m^2(\mathbf{x})d\mathbf{x} \\ L_D(\xi) &= \frac{(\nu + \int_S m^2(\mathbf{x})d\mathbf{x} - 1)}{\gamma^2\gamma_{12}} \\ L_A(\xi) &= \nu(1 + \frac{2}{\gamma} + \frac{1}{\gamma_{12}}) + \int_S m^2(\mathbf{x})d\mathbf{x} - 1 \end{aligned}$$

over $\gamma, \gamma_{12}$. For Q-optimality, we obtain the following system of equations in terms of $\gamma, \gamma_{12}$ and $\nu$,

$$34,560\gamma_{12}^3 + 240(1 - 24\gamma)\gamma_{12}^2 + 240\gamma^2\gamma_{12} + \gamma^2(3 - 56\gamma) = 0,$$
$$270\gamma^2(56\gamma - 240\gamma_{12} - 3) = \nu$$

Likewise for D-optimality, the following system of equations determine the minimizer of $L_D(\xi)$ over $\gamma, \gamma_{12}$,

$$2880\gamma_{12}^2 + (20 - 240\gamma)\gamma_{12} + 3\gamma - 56\gamma^2 = 0,$$
$$2700\gamma_{12}^2\gamma - 225\gamma_{12} + \frac{405}{4}\gamma - 630\gamma^2 - \frac{65}{16} = \nu.$$

In the same vein, the following system of equations determine the minimizer of $L_A(\xi)$ over $\gamma, \gamma_{12}$,

$$2880\gamma_{12}^3 + (20 - 480\gamma)\gamma_{12}^2 + 240\gamma^2\gamma_{12} - 56\gamma^3 + 3\gamma^2 = 0,$$
$$\frac{45}{2}\gamma^2(56\gamma - 240\gamma_{12} - 3) = \nu.$$

Thus, given $\gamma$ we use the above system of equations to determine $\gamma_{12}$, $\nu$, and the coefficients $a, b$, and $c$ are given by (3.10) - (3.12). We have assumed that

$a, b$, and $c$ are nonnegative in this section, consequently we are only able to give solution for small values of $\nu$. The range of values of $\nu$ for which we have nonnegative coefficients for Q-optimality, D-optimality and A-optimality are $0 \leq \nu \leq 1.66$, $0 \leq \nu \leq 0.42$ and $0 \leq \nu \leq 0.86$ respectively. So far all results have hinged on the assumption that the first eigenvalue, is maximum. Numerical verification of this assumption reveals that Q-optimality results are valid in the interval $.21 < \nu \leq 1.66$, whereas in the interval $0 < \nu \leq .21$, the maximum eigenvalue is the second, $\frac{1}{12\gamma^2} \int_S x_1^2 m^2(\mathbf{x}) d\mathbf{x}$ when evaluated at the $m_0$ that minimizes the first one. For this interval we change the density to that corresponding to second eigenvalue. On verifying the results, however, the maximum eigenvalue turns out to be the first when evaluated at the $m_0$ that minimizes the second one. This difficulty motivated us to attempt the generic density model (3.7) reported by Shi (2002) which has the same form irrespective of the maximum eigenvalue. However, owing to the over-parameterization of the model it is numerically intractable. The model seems efficient for low-dimensional problems only. Nevertheless, the difference between this maximum eigenvalue and the other is small. Thus, the results obtained using $\int_S m^2(\mathbf{x}) d\mathbf{x}$ can as well be taken as approximation of the true solution. Figure 1 shows loss plots for some values of $\nu$ in the neighbourhood of the corresponding optimal coefficients. We note that all the plots are very flat in the neighbourhood of these optimal coefficients. This flatness may pose numerical difficulties especially when using gradient based optimization routines.

For D-optimality, the assumption is valid in the interval $0.22 < \nu \leq 0.42$. For the interval $0 < \nu \leq .22$, the maximum eigenvalue is the second eigenvalue, $\frac{1}{\gamma} \int_S x_1^2 m^2(\mathbf{x}) d\mathbf{x} - 12\gamma$ rather than $\int_S m^2(\mathbf{x}) d\mathbf{x} - 1$. Here

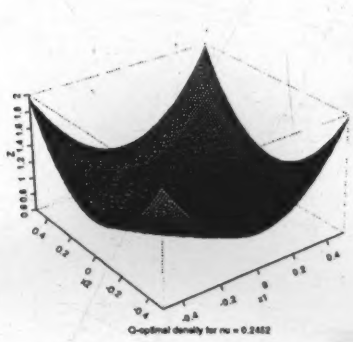
Table 1. Numerical values for Q-optimal minimax densities[1]

| $\nu$ | $\gamma$ | $\gamma_{12}$ | $a$ | $b$ | $c$ | minimum loss |
|---|---|---|---|---|---|---|
| 0 | $\frac{15}{180}$ | 0.0069 | 1 | 0 | 0 | 1 |
| 0.1026 | $\frac{16}{180}$ | 0.0080 | 0.8716 | 0.5409 | 5.5091 | 1.3955 |
| 0.2452 | $\frac{17}{180}$ | 0.0091 | 0.7379 | 1.1455 | 10.2536 | 1.9156 |
| 0.4307 | $\frac{18}{180}$ | 0.0102 | 0.6004 | 1.7947 | 14.4638 | 2.5492 |
| 0.6621 | $\frac{19}{180}$ | 0.0112 | 0.4603 | 2.4761 | 18.2690 | 3.3058 |
| 0.9427 | $\frac{20}{180}$ | 0.0122 | 0.3182 | 3.1816 | 21.8211 | 4.1837 |
| 1.2758 | $\frac{21}{180}$ | 0.0133 | 0.1745 | 3.9055 | 25.1343 | 5.1852 |
| 1.6649 | $\frac{22}{180}$ | 0.0143 | 0.0297 | 4.6438 | 28.2749 | 6.3129 |

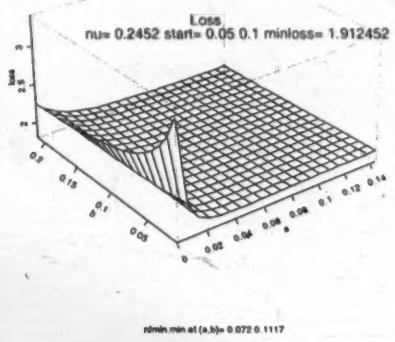
1. $m_0(x) = (a + b(x_1^2 + x_2^2) + cx_1^2x_2^2)^+$

again, we attempt the generic model (3.7) of Shi (2002) for the interval $0 < \nu \leq .22$ but it is not helpful either. Just like the case of Q-optimality the difference between the two eigenvalues is not enough to bother about. The results for A-optimality are completely off the mark. Rather than the first eigenvalue, $\int_S m^2(\mathbf{x})d\mathbf{x} - 1$, being the maximum, the third eigenvalue, $\frac{1}{\gamma_{12}^2}\int_S x_1^2x_2^2m^2(\mathbf{x})d\mathbf{x} - 144$ is maximum.

Table 2 shows the results for D-optimality. We have taken the fourth root of the minimum loss to reflect the actual loss corresponding to $\nu$. This is necessary because the loss is a determinant of a $4 \times 4$ matrix with each entry having a $\nu$ component. Corresponding density and loss plots are shown in Figure 2 for some $\nu$. Again, we note the flatness of the loss plots in the neighbourhood of the optimal coefficients. Going forward we shall treat Q- and D-optimality hand in hand while A-optimality will be considered separately in a later section.

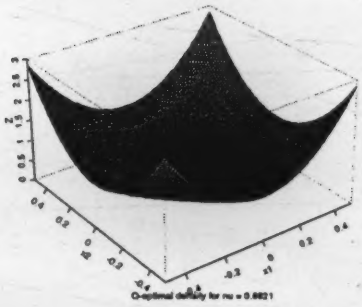


Optimal design density

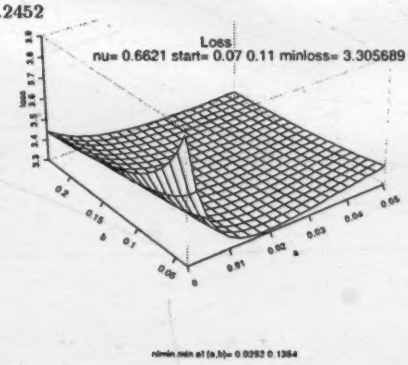



Loss plot

$\nu$=0.2452

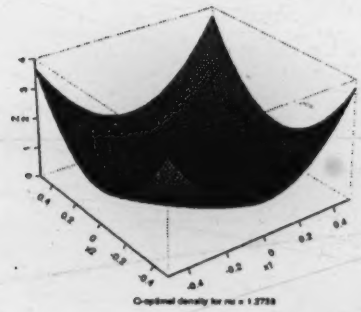


Optimal design density

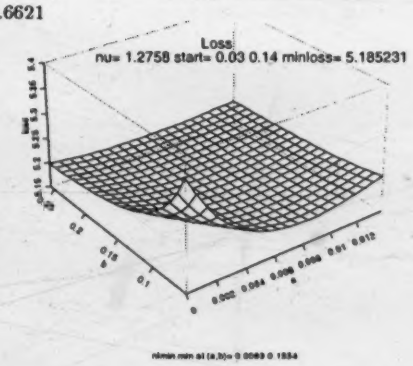



Loss plot

$\nu$=0.6621



Optimal design density




Loss plot

$\nu$=1.2758

Figure 1. Q-optimal design densities and loss plots for some $\nu$

Table 2. Numerical values for D-optimal minimax densities[1]

| $\nu$ | $\gamma$ | $\gamma_{12}$ | $a$ | $b$ | $c$ | (minimum loss)$^{1/4}$ |
|---|---|---|---|---|---|---|
| 0 | $\frac{15}{180}$ | 0.0069 | 1 | 0 | 0 | 0 |
| 0.0804 | $\frac{16}{180}$ | 0.0080 | 0.8733 | 0.5207 | 5.7514 | 6.1770 |
| 0.1547 | $\frac{17}{180}$ | 0.0091 | 0.7438 | 1.0746 | 11.1050 | 7.0640 |
| 0.2225 | $\frac{18}{180}$ | 0.0102 | 0.6123 | 1.6526 | 16.1685 | 7.5409 |
| 0.2834 | $\frac{19}{180}$ | 0.0113 | 0.4793 | 2.2489 | 21.0134 | 7.8368 |
| 0.3372 | $\frac{20}{180}$ | 0.0124 | 0.3451 | 2.8593 | 25.6890 | 8.0319 |
| 0.3840 | $\frac{21}{180}$ | 0.0134 | 0..2099 | 3.4808 | 30.2307 | 8.1639 |
| 0.4235 | $\frac{22}{180}$ | 0.0144 | 0.0741 | 4.1113 | 34.6645 | 8.2538 |

1. $m_0(x) = (a + b(x_1^2 + x_2^2) + cx_1^2x_2^2)^+$

## 3.5 Larger Values of $\nu$ for Q- and D-optimality

Here we still want to hold to the assumption that

$$ch_{max}(\mathbf{K}(\xi)\mathbf{H}^{-1}(\xi)) = \int_S m^2(\mathbf{x})d\mathbf{x},$$
$$ch_{max}(\mathbf{G}(\xi)\mathbf{B}^{-1}(\xi)) = \int_S m^2(\mathbf{x})d\mathbf{x} - 1.$$

For larger values of $\nu$ corresponding to

$$\gamma > \frac{22}{180},$$

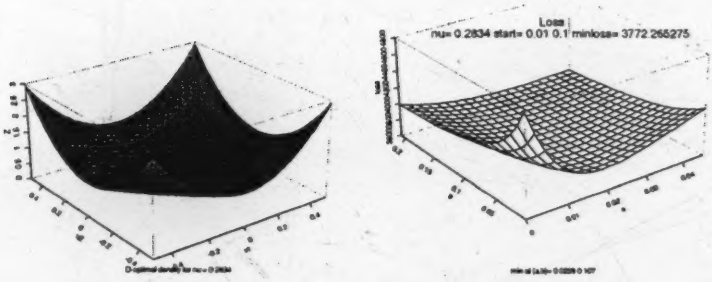
our findings reveal that $a, b$ and $c$ cannot all be nonnegative. Thus the density

$$m_0(\mathbf{x}) = (a + b(x_1^2 + x_2^2) + cx_1^2x_2^2)^+$$

cannot be simplified further. As a result our approach in seeking the optimal design will be exclusively numerical.

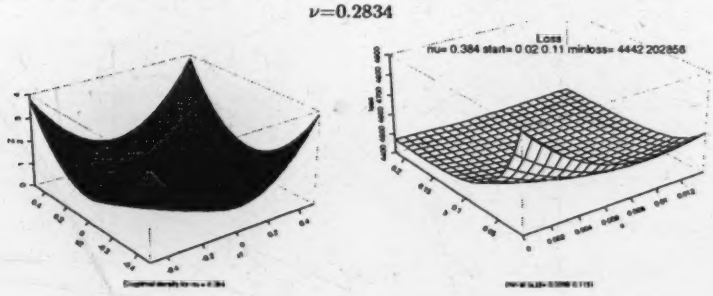
We aim, as in for small values of $\nu$, to minimize

$$L_Q(\xi) = \nu(1 + \frac{1}{6\gamma} + \frac{1}{144\gamma_{12}}) + \int_S m_0^2(\mathbf{x})d\mathbf{x}$$
$$L_D(\xi) = (\nu + \int_S m_0^2(\mathbf{x})d\mathbf{x} - \mathbf{1})/\gamma^2\gamma_{12}$$



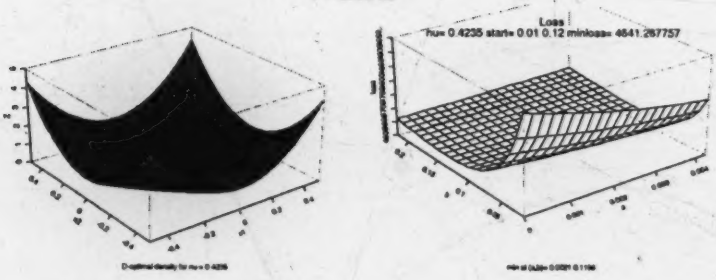

Optimal design density Loss plot

$\nu$=0.2834




Optimal design density Loss plot

$\nu$=0.3840




Optimal design density Loss plot

$\nu$=0.4235

Figure 2. D-optimal design densities and loss plots for some $\nu$

with respect to $a, b$ and $c$ subject to (3.8). Since we would be solving this problem numerically it is appropriate to restate it as follows,

minimize

$$\begin{aligned} L_Q(\xi) &= \nu\left(1+\frac{1}{6\int_S x_1^2 m_0(\mathbf{x})\,d\mathbf{x}}+\frac{1}{144\int_S x_1^2x_2^2 m_0(\mathbf{x})\,d\mathbf{x}}\right) \\ &\quad +\int_S m_0^2(\mathbf{x})d\mathbf{x}, \\ L_D(\xi) &= \frac{\left(\nu+\int_S m_0^2(\mathbf{x})\,d\mathbf{x}-1\right)}{\left(\int_S x_1^2 m_0(\mathbf{x})\,d\mathbf{x}\right)^2\left(\int_S x_1^2x_2^2 m_0(\mathbf{x})\,d\mathbf{x}\right)}, \end{aligned}$$

subject to

$$\int_S m_0(\mathbf{x})d\mathbf{x}=1.$$

Our earlier attempt was to use an optimization routine for nonlinear optimization with constraints, "fmincon" in MATLAB. This routine is a gradient based algorithm and hence its failure because of the flatness of the loss functions in the neighbourhoods of the optimal coefficients. At this point we conclude that for the kind of problems we are attempting to proffer solutions, gradient based methods may not be viable since most of these methods assume the existence of first and second-order partial derivatives of the objective function at every point in its domain. The integrals indicated in the loss functions and the density condition are actually double integrals (with respect to both $x_1$ and $x_2$) and the have been evaluated using "dblquad", a numerical double integration routine in MATLAB. The "dblquad" routine implements a low order method using an adaptive recursive Simpson's rule (See Gander and Gautschi 2000). Our next attempt was to use an optimization routine for nonlinear optimization without constraints, "fminsearch" also in MATLAB. This routine is based on Nelder-Mead Simplex Algorithm which is a direct search method and does not require the evaluation of any

partial derivatives of the objective function. To adapt this problem to this optimization routine, we built the density constraint

$$\int_S m_0(\mathbf{x})d\mathbf{x} = 1,$$

into the objective function such that the optimization is carried out on parameters $a$ and $b$ while $c$ is chosen by a solution to

$$g(c) \equiv \int_S m_0(\mathbf{x})d\mathbf{x} - 1 = 0$$

using Newton-Raphson's algorithm. This approach worked well for small values of $\nu$ for which we have presented the results in the last section. However, the results for larger values of $\nu$ are not acceptable, in the sense that the "optimal" values gotten for $a, b$ and $c$ failed to satisfy the density condition.

As a further step to rectify this problem, we attempt using the density

$$m_0(\mathbf{x}) = \frac{(a' + b'(x_1^2 + x_2^2) + c'x_1^2x_2^2)^+}{\int_S (a' + b'(x_1^2 + x_2^2) + c'x_1^2x_2^2)^+ d\mathbf{x}}.$$

Using this model we expect that for any value of $a', b'$ and $c'$ the condition $\int_S m_0(\mathbf{x})d\mathbf{x} = 1$ should be satisfied. In the parameterization of the original density

$$\begin{aligned} a &= \frac{a'}{\int_S (a' + b'(x_1^2 + x_2^2) + c'x_1^2x_2^2)^+ d\mathbf{x}}, \\ b &= \frac{b'}{\int_S (a' + b'(x_1^2 + x_2^2) + c'x_1^2x_2^2)^+ d\mathbf{x}}, \\ c &= \frac{c'}{\int_S (a' + b'(x_1^2 + x_2^2) + c'x_1^2x_2^2)^+ d\mathbf{x}}. \end{aligned}$$

With this approach we encounter the same problem as with the first approach, the density condition is violated for the supposed optimal parameters for large values of $\nu$. Worse still, this approach is numerically inefficient.

We further refine our model by dividing the numerator and denominator by $a'$ so as to reduce its complexity. Thus, we have

$$m_0(\mathbf{x}) = \frac{(1 + a'(x_1^2 + x_2^2) + b'x_1^2x_2^2)^+}{\int_S (1 + a'(x_1^2 + x_2^2) + b'x_1^2x_2^2)^+ d\mathbf{x}}.$$

In the parameterization of the original density

$$\begin{aligned} a &= \frac{1}{\int_S (1 + a'(x_1^2 + x_2^2) + b'x_1^2x_2^2)^+ d(\mathbf{x})} \\ b &= \frac{a'}{\int_S (1 + a'(x_1^2 + x_2^2) + b'x_1^2x_2^2)^+ d(\mathbf{x})}, \\ c &= \frac{b'}{\int_S (1 + a'(x_1^2 + x_2^2) + b'x_1^2x_2^2)^+ d(\mathbf{x})}. \end{aligned}$$

While this approach is faster because of the reduction in the complexity of the problem, it also failed to meet the density condition for large values of $\nu$. However, looking at the pattern of the results obtained we realized in the parameterization of the original density the parameter $a$ is constantly assuming the value zero. Thus we observed that the optimal value may actually be negative but our model definition is not giving it the freedom to assume negative values.

At this juncture, we resort to the nonlinear unconstrained optimization routine, "nlmin" in S-plus with slight modification in the form of the density, for reason stated above. Here we use

$$m_0(\mathbf{x}) = \frac{(a' + b'(x_1^2 + x_2^2) + x_1^2x_2^2)^+}{\int_S (a' + b'(x_1^2 + x_2^2) + x_1^2x_2^2)^+ d\mathbf{x}}.$$

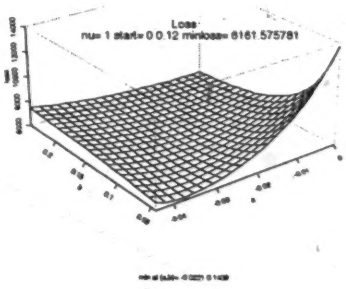
Table 3. Numerical values for Q-optimal minimax densities[1] (Large $\nu$)

| $\nu$ | $\gamma$ | $\gamma_{12}$ | $a$ | $b$ | $c$ | minmum loss |
|---|---|---|---|---|---|---|
| 2 | 0.1263 | 0.0151 | −0.0848 | 5.2217 | 30.5848 | 7.2557 |
| 5 | 0.1456 | 0.0196 | −1.1591 | 9.8268 | 45.3675 | 15.0883 |
| 10 | 0.1591 | 0.0237 | −2.9763 | 16.4536 | 61.7779 | 27.1275 |
| 100 | 0.1981 | 0.0387 | −36.6982 | 106.1545 | 231.3685 | 216.3778 |
| 500 | 0.2170 | 0.0469 | −190.6043 | 442.2879 | 789.6082 | 996.4184 |
| 1000 | 0.2231 | 0.0496 | −384.2855 | 836.9180 | 1409.0691 | 1945.9953 |

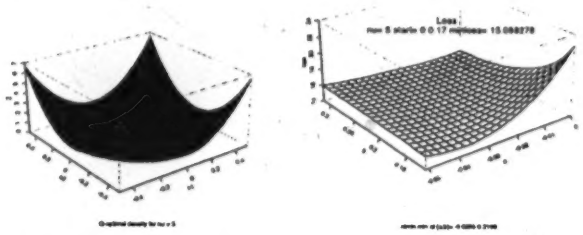
1. $m_0(x) = (a + b(x_1^2 + x_2^2) + cx_1^2x_2^2)^+$

This routine is faster than previous attempts and the optimal results using this routine satisfy the density condition perfectly and as expected. In evaluating the double integrals invloved in the computation we have written S-plus code based on Simpson's rule. Q-optimal results for some values of $\nu$ are outlined in Table 3. The assumption that the maximum eigenvalue is $\int_S m^2(x)dx$ is valid for all $\nu$, thus we made a correct initial guess.

In the case of D-optimality the first eigenvalue remains the maximum eigenvalue for values of $\nu < 400$. Results of some $\nu$ within this range are shown in Table 4.

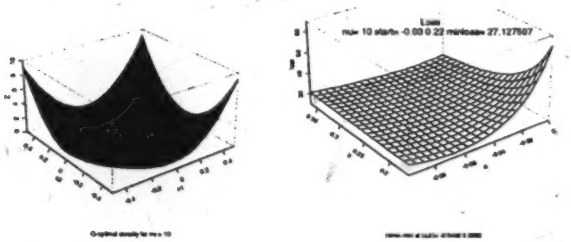



Loss plot



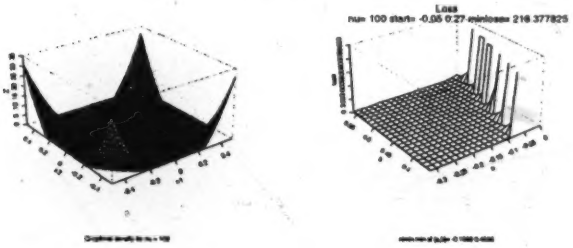
Optimal design density Loss plot

$\nu$=5



Optimal design density Loss plot

$\nu$=10



Optimal design density Loss plot

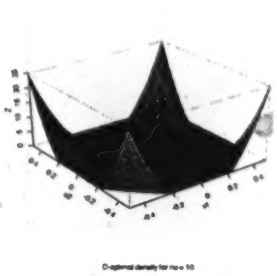
$\nu$=100

Figure 3. Q-optimal design densities and loss plots for some large $\nu$

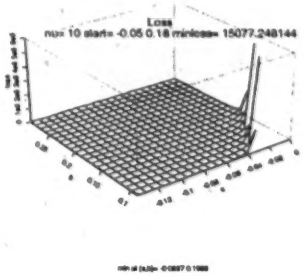
Table 4. Numerical values for D-optimal minimax densities[1] ($\nu < 400$)

| $\nu$ | $\gamma$ | $\gamma_{12}$ | $a$ | $b$ | $c$ | (minmum loss)$^{1/4}$ |
|---|---|---|---|---|---|---|
| 1 | 0.1533 | 0.0221 | −1.5985 | 10.4242 | 72.4443 | 8.8598 |
| 5 | 0.1821 | 0.0324 | −9.8881 | 32.3295 | 181.6528 | 10.2301 |
| 10 | 0.1918 | 0.0362 | −19.0371 | 52.3399 | 277.2301 | 11.0810 |
| 100 | 0.2176 | 0.0471 | −168.3761 | 318.8922 | 1471.8033 | 15.7920 |
| 250 | 0.2249 | 0.0505 | −407.6055 | 705.2664 | 3143.4546 | 18.7770 |
| 350 | 0.2272 | 0.0515 | −565.2849 | 951.7824 | 4200.0021 | 20.0806 |

1. $m_0(x) = (a + b(x_1^2 + x_2^2) + cx_1^2x_2^2)^+$

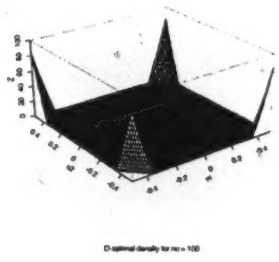


Optimal design density

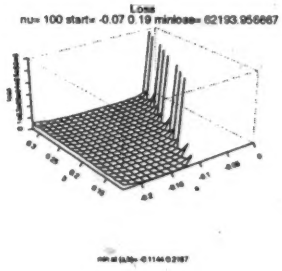


Loss plot

$\nu$=10



Optimal design density



Loss plot

$\nu$=100

For values of $\nu > 400$, the second eigenvalue, $\frac{1}{\gamma}\int_S x_1^2 m^2(\mathbf{x})d\mathbf{x} - 12\gamma$ is maximum. Thus the optimal density is no longer of the form

$$m_0(x) = (a + b(x_1^2 + x_2^2) + cx_1^2x_2^2)^+$$

but

$$m_0(x) = \frac{(a + b(x_1^2 + x_2^2) + cx_1^2x_2^2)^+}{x_1^2 + x_2^2}.$$

For the purpose of numerical computation we have used the form

$$m_0(x) = \frac{(a' + b'(x_1^2 + x_2^2) + x_1^2x_2^2)^+/(x_1^2 + x_2^2)}{\int_S (a' + b'(x_1^2 + x_2^2) + x_1^2x_2^2)^+/(x_1^2 + x_2^2)d\mathbf{x}}$$

in order to reduce computational complexity and to ensure that the density condition $\int_S m(\mathbf{x})d\mathbf{x} = 1$ is satisfied. In the parameterization of the original

Table 5. Numerical values for D-optimal minimax densities[1] ($\nu > 400$)

| $\nu$ | $\gamma$ | $\gamma_{12}$ | $a$ | $b$ | $c$ | (minmum loss)$^{1/4}$ |
|---|---|---|---|---|---|---|
| 500 | 0.2295 | 0.0526 | −391.2978 | 656.7628 | 2630.6546 | 21.6020 |
| 600 | 0.2307 | 0.0532 | −469.4835 | 778.8154 | 3114.0468 | 22.4381 |
| 750 | 0.2318 | 0.0537 | −582.9923 | 951.4849 | 3811.7216 | 23.5193 |
| 1000 | 0.2334 | 0.0544 | −774.7152 | 1242.7007 | 4972.3585 | 25.0120 |

1. $m_0(x) = \frac{(a+b(x_1^2+x_2^2)+cx_1^2x_2^2)^+}{x_1^2+x_2^2}$

density

$$
\begin{aligned}
a &= \frac{a'}{\int_S (a' + b'(x_1^2+x_2^2) + x_1^2x_2^2)^+/(x_1^2+x_2^2)d\mathbf{x}} \\
b &= \frac{b'}{\int_S (a' + b'(x_1^2+x_2^2) + x_1^2x_2^2)^+/(x_1^2+x_2^2)d\mathbf{x}}, \\
c &= \frac{1}{\int_S (a' + b'(x_1^2+x_2^2) + x_1^2x_2^2)^+/(x_1^2+x_2^2)d\mathbf{x}}.
\end{aligned}
$$

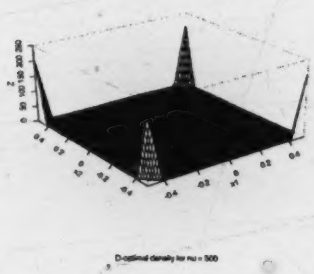
Table 5 presents results for some $\nu$ within this range.

## 3.6 Optimal Coefficients for A-Optimality

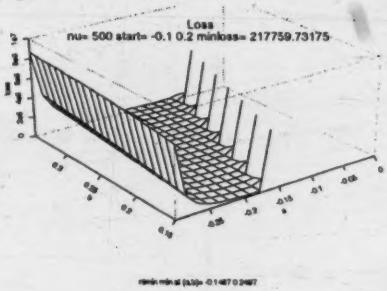
Our attempt in section 3.4 was based on the assumption that the maximum eigenvalue is $\int_S m^2(x)d\mathbf{x}$ but this assumption fails for the supposed optimal coefficients, the third eigenvalue, $\frac{1}{\gamma_{12}^2}\int_S x_1^2x_2^2m^2(x)d\mathbf{x} - 144$ being the maximum.

Nevertheless, this attempt left us with a clue. In our search for solutions for A-optimality we shall start by assuming that the third eigenvalue is maximum, that is

$$ch_{max}(\mathbf{G}(\boldsymbol{\xi})\mathbf{B}^{-2}(\boldsymbol{\xi})) = \frac{1}{\gamma_{12}^2}\int_S x_1^2x_2^2m^2(\mathbf{x})d\mathbf{x} - 144.$$

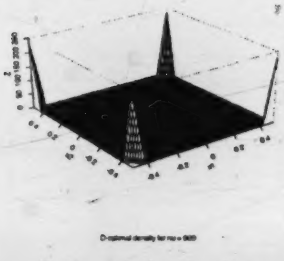


Optimal design density

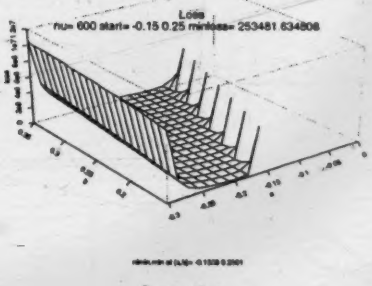



Loss plot

$\nu$=500

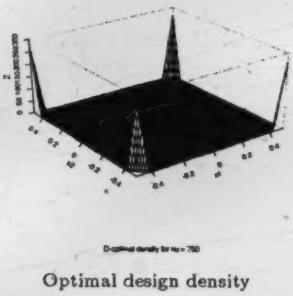


Optimal design density

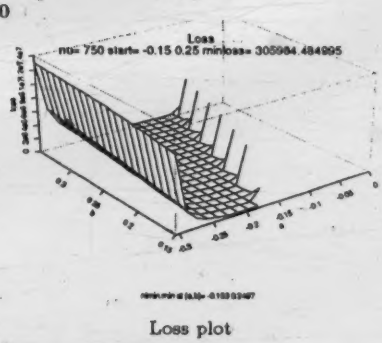



Loss plot

$\nu$=600

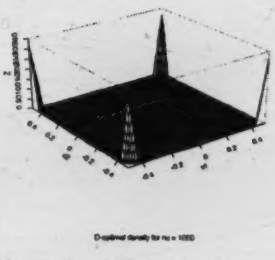


Optimal design density

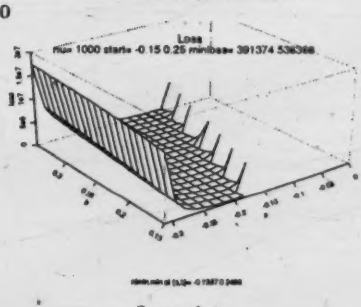



Loss plot

$\nu$=750



Optimal design density




Loss plot

$\nu$=1000

Figure 5. D-optimal design densities and loss plots for some $\nu$ ($\nu$>400)

Table 6. Numerical values for A-optimal minimax densities[1]

| $\nu$ | $\gamma$ | $\gamma_{12}$ | $a$ | $b$ | $c$ | minimum loss |
|---|---|---|---|---|---|---|
| 0.2 | 0.1165 | 0.0125 | -0.0015 | 0.0012 | 1.9438 | 22.6511 |
| 0.5 | 0.1279 | 0.0152 | -0.0047 | 0.0035 | 2.5699 | 49.3128 |
| 1 | 0.1382 | 0.0180 | -0.0113 | 0.0085 | 3.3807 | 87.2778 |
| 5 | 0.1656 | 0.0265 | -0.0875 | 0.0641 | 8.0107 | 317.6984 |
| 10 | 0.1775 | 0.0308 | -0.2072 | 0.1502 | 12.7078 | 554.2387 |
| 100 | 0.2103 | 0.0440 | -3.1183 | 2.1859 | 77.7624 | 3812.1171 |
| 500 | 0.2255 | 0.0507 | -18.3982 | 12.6665 | 324.9430 | 16120.3951 |
| 1000 | 0.2303 | 0.0530 | -38.6586 | 26.4433 | 616.6707 | 30626.3085 |

1. $m_0(x) = \frac{(a+b(x_1^2+x_2^2)+cx_1^2x_2^2)^+}{x_1^2x_2^2}$

Therefore, the optimal density is of the form

$$m_0(\mathbf{x}) = \frac{(a + b(x_1^2 + x_2^2) + cx_1^2x_2^2)^+}{x_1^2x_2^2}.$$

Thus, given that $m_0(x)$ is defined as above we minimize

$$L_A(\xi) = \nu\left(1 + \frac{2}{\int_S x_1^2 m_0(\mathbf{x})d\mathbf{x}} + \frac{1}{\int_S x_1^2x_2^2 m_0(\mathbf{x})d\mathbf{x}}\right) + \frac{\int_S x_1^2x_2^2 m_0^2(\mathbf{x})d\mathbf{x}}{\left(\int_S x_1^2x_2^2 m_0(\mathbf{x})d\mathbf{x}\right)^2} - 144$$

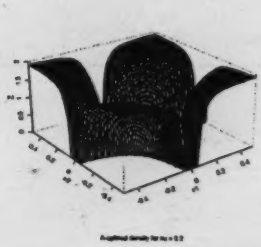
subject to

$$\int_S m_0(\mathbf{x})d\mathbf{x} = 1.$$

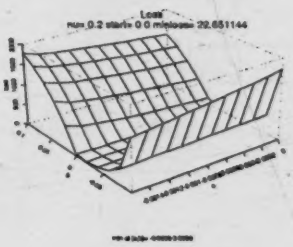
Here again, we use

$$m_0(\mathbf{x}) = \frac{(a' + b'(x_1^2 + x_2^2) + x_1^2x_2^2)^+/x_1^2x_2^2}{\int_S (a' + b'(x_1^2 + x_2^2) + x_1^2x_2^2)^+/x_1^2x_2^2\, d\mathbf{x}}.$$

for computational purpose. Table 6 gives the optimal coefficients for some values of $\nu$ in the parameterization of the original density.

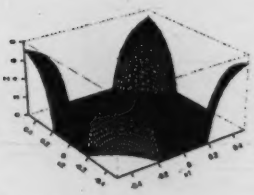


Optimal design density

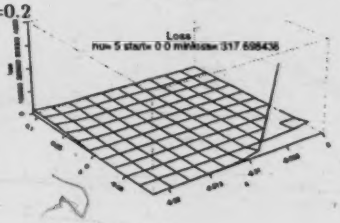


Loss plot

$\nu$=0.2

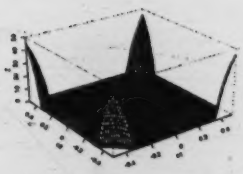


Optimal design density

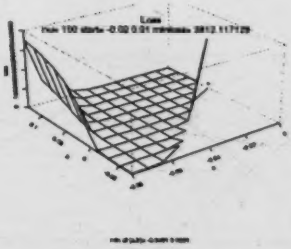


Loss plot

$\nu$=5

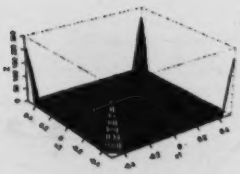


Optimal design density

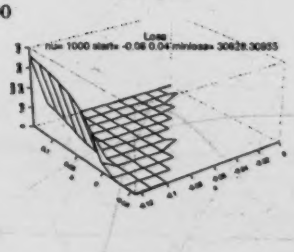


Loss plot

$\nu$=100



Optimal design density



Loss plot

$\nu$=1000

Figure 6. A-optimal design densities and loss plots for some $\nu$

Explicit expressions, obtained for $\nu = 10$ are

Q-optimality:

$$m(x_1, x_2) = (-2.9763 + 16.4536(x_1 + x_2) + 61.7779x_1x_2)^+,$$

D-optimality:

$$m(x_1, x_2) = (-19.0371 + 52.3399(x_1 + x_2) + 277.2301x_1x_2)^+, \text{ and}$$

A-optimality:

$$m(x_1, x_2) = (-0.2072 + 0.1502(x_1 + x_2) + 12.7078x_1x_2)^+/x_1^2x_2^2$$

Explicit expressions,obtained for $\nu = 100$ are

Q-optimality:

$$m(x_1, x_2) = (-36.6982 + 106.1545(x_1 + x_2) + 231.3685x_1x_2)^+$$

D-optimality:

$$m(x_1, x_2) = (-168.3761 + 318.8922(x_1 + x_2) + 1471.8033x_1x_2)^+$$

A-optimality:

$$m(x_1, x_2) = (-3.1183 + 2.1859(x_1 + x_2) + 77.7624x_1x_2)^+/x_1^2x_2^2$$

Explicit expressions,obtained for $\nu = 1000$ are

Q-optimality:

$$m(x_1, x_2) = (-384.2855 + 836.9180(x_1 + x_2) + 1409.0691x_1x_2)^+,$$

D-optimality:

$$m(x_1, x_2) = (-774.7152 + 1242.7007(x_1 + x_2) + 4972.3585x_1x_2)^+/(x_1^2 + x_2^2),$$

and

A-optimality:

$$m(x_1, x_2) = (-38.6586 + 26.4433(x_1 + x_2) + 616.6707x_1x_2)^+ / x_1^2 x_2^2$$

# Chapter 4

# IMPLEMENTATIONS

In this chapter, we implement some of the optimal densities derived in the last chapter. The implementation of a continuous measure as a discrete design involves some arbitrariness. However, this implementation is intuitively sensible as well as robust. A balance is struck between full efficiency and robustness as we place observations at varied locations near the sites at which the variance-minimizing designs place all of their mass. We implement our designs using two case studies, waste water ozonation problem reported by Heo, Schmuland and Wiens (2001) and an engineering assembly problem.

## 4.1 Case Study 1: Waste Water Ozonation

Prairie farmers in Alberta have traditionally stocked dugouts with trout for recreational purposes. Some are now attempting commercial fish culturing indoors, year-round. Because of limited water supplies, attempts are being made to recycle waste water for this purpose. Most solids in waste water from trout-rearing facilities settle readily, but a suspension of fine "particulate" material remains. Several studies have shown that fine particulate adversely affects fish health and productivity. The wastewater engineering

research team at the Alberta Environment Centre conducted a bench-scale experiment to determine the amount of total suspended solid (TSS) remaining after applying ozone ($O_3$) at application rates ranging from 0 to 2 mg/L (Heo and James 1995). Because ozonation is to be used for disinfection and the associated capital cost is high, the team wanted to determine an optimal $O_3$ rate, minimizing the worst cost. Another factor which is important in the removal of suspended solids is the gas to liquid ratio, denoted GL. Uncertainties about the exact nature of the relationship between TSS, $O_3$ and GL led to the assumption of an approximate partial second order model with interaction. That is,

$$TSS = \theta_0 + \theta_1(O_3) + \theta_2(GL) + \theta_{12}(O_3GL) + error. \tag{4.1}$$

Both factors were recoded linearly such that $O_3, GL \in [-1/2, 1/2]$. The Q-optimal design $\xi$, with $\nu = 5$ , was then implemented as follows to yield $n = 48$ design points. We choose $n_0 = n/8$ points $(x_1, x_2)$ in $\{0 \leq x_1 \leq x_2 \leq 1/2\}$ and then obtained the remaining $7n_0$ sites by symmetry and exchangeability. The $n_0$ points $\{x_{1i}, x_{2i}\}_{i=1}^{n_0}$ were chosen such that the moments

$$e_{2j,2k} = \sum_{i=1}^{n_0} (x_{1i}^{2j} x_{2i}^{2k} + x_{1i}^{2k} x_{2i}^{2j})/(2n_0) \tag{4.2}$$

matched up as closely as possible with the theoretical moments $E_\xi(X_1^{2j} X_2^{2k})$ obtained from $\xi$. We did this for the $J(J+3)/2$ choices $(k, j)$ with $k = 0, ..., j$ and $j = 1, ..., J$, with $J$ being the smallest integer for which $J(J+3)/2$ exceeds the number $2n_0$ of coordinates to be chosen. Thus $n = 48, n_0 = 6$ yielded $J = 4$ and 14 even order moments to be matched up. Of course, all moments with at least one odd order are zero, and the 14 moments obtained by exchanging $j$ and $k$ will be matched as well. The matching was done by

numerical minimization of

$$\sum_{j,k}\{e_{2j,2k} - E_\xi(X_1^{2j}X_2^{2k})\}^2, \tag{4.3}$$

yielding the implementation shown in Figure 7 with

$$\begin{aligned}\{x_{1i}, x_{2i}\}_{i=1}^{n_0} &= \{(0.036, 0.422), (0.372, 0.272), (0.409, 0.433),\\ &\quad (0.493, 0.243), (0.315, 0.417), (0.42, 0.5)\}.\end{aligned}$$

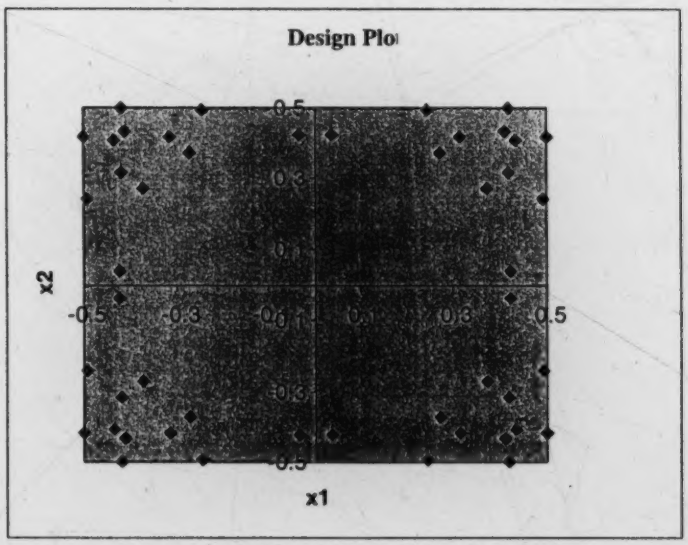



Figure 7. Implementation of the Q-optimal design at $\nu$=5

## 4.2 Case Study 2: Injection Molding Process

In an engineering assembly, parts manufactured in an injection molding process are showing excessive shrinkage. This is causing problems in assembly operations downstream from the injection molding area. A quality improvement team has decided to use a regression model to study the injection

molding process with a focus on reducing the shrinkage. The team investigated factors affecting the shrinkage and decided that two factors - mold temperature (T) and screw speed (S) are key. The associated heating cost is high, so the team is interested in determining an optimal mold temperature level, minimizing the worst cost. Uncertainties about the exact nature of the relationship between $SH, T$ and $S$ and previous assembly experience led to the assumption of an approximate partial second order model.

Both factors were linearly transformed to the range [-1/2,1/2]. The A-optimal design with $\nu = 100$, was implemented as described in the last section. The result of the implementation is

$$\begin{aligned}\{x_{1i}, x_{2i}\}_{i=1}^{n_0} &= \{(0.445, 0.464), (0.465, 0.423), (0.489, 0.489) \\ &\quad (0.466, 0.399), (0.495, 0.432), (0.427, 0.497)\}.\end{aligned}$$

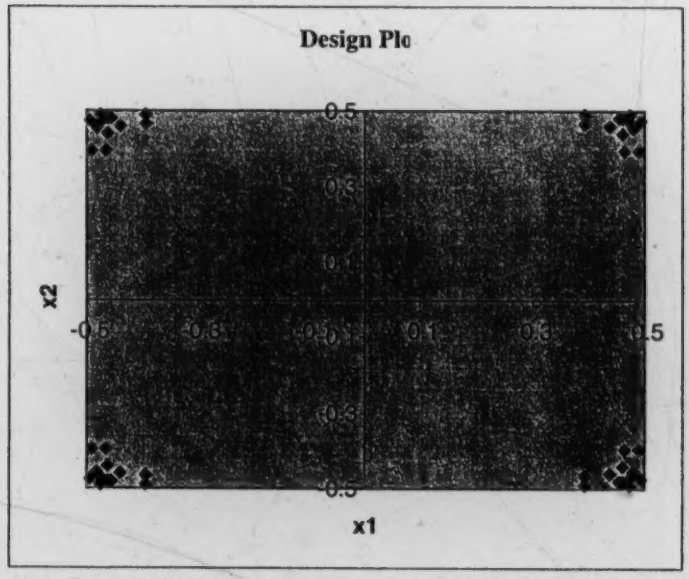



Figure 8. Implementation of the A-optimal design at $\nu$=100

# REFERENCES


Box, G.E.P., and Draper, N.R. (1959), "A basis for the selection of a response surface design", Journal of American Statistical Association, 54, 622 - 654.

Gander, W. and Gautschi W. (2000), "Adaptive Quadrature - Revisited", BIT, 40, 84 - 107.

Heo, G. and James, W. (1995), "The Effects of Ozone Application Rate and Gas:Liquid Ratio on Removing Fine Suspended Particles in a Bubble Column Reactor," Technical report, Alberta Environmental Centre.

Heo, G., Schmuland, B. and Wiens, D.P. (1999), "Restricted Minimax Robust Designs for Misspecified Regression Models," Technical report 99.04, Statistics Center, Department of Mathematical Sciences, University of Alberta, Edmonton, Alberta, Canada.

Heo, G., Schmuland, B. and Wiens, D.P. (2001), "Restricted Minimax Robust Designs for Misspecified Regression Models," The Canadian Journal of Statistics, 29, 117-128.

Huber, P.J. (1975), "Robustness and Designs," in: A Survey of Statistical Design and Linear Models, ed. J.N. Srivastava, Amsterdam: North Holland, pp. 287-303.

Lagarias, J.C., Reeds, J.A., Wright, M.H. and Wright, P.E. (1998), "Convergence Properties of Nelder-Mead Simplex Method in Low Dimension," SIAM Journal of Optimization 9, 112 - 147.

Li, K.C., and Notz, W. (1982), "Robust Designs for Nearly Linear Regression," Journal of Statistical Planning and Inference, 6, 135-151.

Liu, S.X., and Wiens, D.P. (1997), "Robust Designs for Approximately Polynomial Regression," Journal of Statistical Planning and Inference, 64, 369-381.

Marcus, M.B. and J. Sacks (1976, "Robust Designs for Regression Problems," in: S.S Gupta and D.S. Moore, Eds., Statistical Theory and Related Topics II. Academics Press, New York, 245 - 268.

Pesotchinsky, L. (1982), "Optimal Robust Designs: Linear Regression in Rk," The Annals of Statistics, 10, 511-525.

Sacks, J., and Ylvisaker, D. (1978), "Linear Estimation for Approximately Linear Models," The Annals of Statistics, 6, 1122-1137.

Shi, P. (2002), "Minimax Robust Designs for Misspecified Regression Models," Unpublished Doctoral Dissertation, University of Victoria, British Columbia, Canada.

Wiens, D.P. (1990), "Robust, Minimax Designs for Multiple Linear Regression," Linear Algebra and Its Applications, Second Special Issue on Linear Algebra and Statistics; 127, 327 - 340.

Wiens, D.P. (1992), "Minimax Designs for Approximately Linear Regression," Journal of Statistical Planning and Inference, 31, 353-371.

# APPENDIX: SAMPLE PROGRAM CODES

The following are examples of S-PLUS codes used for the computations.

```
simp2 <- function(mat,lx = -0.5,ux = 0.5,ly = -0.5,uy = 0.5)

{
#This program evaluates double integrals with appropriate
#limits of integration specified as lx,ux and ly,uy.
#Code written by D.P. Wiens using Simpson's rule

 mx <- nrow(mat) - 1
 hx <- (ux - lx)/mx
 dx <- c(1, rep(c(4, 2), mx/2 - 1), 4, 1)
 my <- ncol(mat) - 1
 hy <- (uy - ly)/my
 dy <- c(1, rep(c(4, 2), my/2 - 1), 4, 1)
 (hx * hy * sum(outer(dx, dy) * mat))/9
}
```

```
pdensity_function(a,b,N) {
#This function evaluates the density function over
#a grid of points determined by x and y

 x_seq(-.5,.5,length=N)
 y_seq(-.5,.5,length=N)
 mat <- (a+b*outer(x^2,y^2,''+'')+outer(x^2,y^2,''*''))
 mat <- pmax(mat,0)
 integral_simp2(mat)
mat/integral
}
```

```
gamm_function(a,b,N) {
#This function evaluates gamma over a grid
#of points determined by x and y

 x_seq(-.5,.5,length=N)
 y_seq(-.5,.5,length=N)
 mat1 <- outer(x^2,rep(1,N))
 mat <- mat1*pdensity(a,b,N)
 simp2(mat)
}
```

```
gam12_function(a,b,N) {
#This function evaluates gamma12 over a
#grid of points determined by x and y

 x_seq(-.5,.5,length=N)
 y_seq(-.5,.5,length=N)
 mat2 <- outer(x^2,y^2)
 mat <- mat2*pdensity(a,b,N)
 simp2(mat)
}

evalue1_function(a,b,N) {
#This function evaluates first eigenvalue(IMSE)
#over a grid of points determined by x and y

 mat <- (pdensity(a,b,N))^2
 simp2(mat)
}

evalue2_function(a,b,N) {
#This function evaluates second eigenvalue(IMSE)
#over a grid of points determined by x and y

 x_seq(-.5,.5,length=N)
 y_seq(-.5,.5,length=N)
 mat3 <- outer(x^2,rep(1,N))
 mat <- mat3*(pdensity(a,b,N))^2
 simp2(mat)/(12*(gamm(a,b,N))^2)
```

```
}

evalue3_function(a,b,N) {
#This function evaluates third eigenvalue(IMSE)
#over a grid of points determined by x and y

 x_seq(-.5,.5,length=N)
 y_seq(-.5,.5,length=N)
 mat4 <- outer(x^2,y^2)
 mat <- mat4*(pdensity(a,b,N))^2
 simp2(mat)/(144*(gam12(a,b,N)^2))
}

loss_function(v,r,N=101) {
#This function evaluates IMSE loss

 a <- v[1]
 b <- v[2]
 e1 <- evalue1(a,b,N)
 e2 <- evalue2(a,b,N)
 e3 <- evalue3(a,b,N)
 g1 <- gamm(a,b,N)
 g2 <- gam12(a,b,N)
 ch.max <- max(e1,e2,e3)
 loss <- r * (1 + 1/(6*g1) + 1/(144*g2)) + e1
 list(loss=loss, ch.max=ch.max, e1=e1)
}
```

```
minimise_function(r,N,start){
#This function minimizes the loss over a and b

 params <- c(r,N)
 assign(''params'',params,frame=1)
 fn <- function(v) loss(v,params[1],params[2])$loss
 qwe <- nlmin(fn,start,max.iter=10000)
 check.loss <- loss(qwe$x,params[1],params[2])
 check <- (check.loss$ch.max==check.loss$e1)
 cat(''check='',check,''\n'')
 a <- qwe$x[1]
 b <- qwe$x[2]
 x_seq(-.5,.5,length=N)
 y_seq(-.5,.5,length=N)
 mat <- (a+b*outer(x^2,y^2,''+'')+outer(x^2,y^2,''*''))
 mat <- pmax(mat,0)
 integral_simp2(mat)
 lmd <- c(a,b,1)/integral
 cat(''lambda, mu, delta ='', round(lmd,4),''\n'')
 qwe
}

plotloss <-function(r, N=101, start, width1, width2, L=11)
{
#Plots the loss in a neighbourhood of
#the minimum and the optimal density
#corresponding to the minimum loss
```

```
qwe <- minimise(r, N, start)
print(qwe)
a <- qwe$x
width1 <- min(width1, 2*a[1])
width2 <- min(width2, 2*a[2])
g1 <- gamm(a[1],a[2],N)
g2 <- gam12(a[1],a[2],N)
e1 <- evalue1(a[1],a[2],N)
e2 <- evalue2(a[1],a[2],N)
e3 <- evalue3(a[1],a[2],N)
minloss <- r * (1 + 1/(6*g1) + 1/(144*g2)) + e1
cat(''loss at (a,b)='',minloss,''\n'')
cat(''gamma at (a,b)='',g1,''\n'')
cat(''gamma12 at (a,b)='',g2,''\n'')
cat(''evalues : '', round(e1,6),round(e2,6),
\qquad round(e3,6),''\n'')
x1 <- seq(a[1]-.5*width1,a[1]+.5*width1,length=L)
x2 <- seq(a[2]-.5*width2,a[2]+.5*width2,length=L)
mat <- matrix(nrow = L, ncol = L)
for(i in 1:L){
for(j in 1:L){
mat[i,j] <- loss(c(x1[i],x2[j]),r,N)$loss
}
}
mat2 <- pdensity(a[1],a[2],N)
par(oma=c(6,0,0,0))
pp <- persp(x1,x2,mat,xlab=''a'',ylab=''b'',zlab=''loss'')
title(main=paste(''Loss'',''\n'',''nu='',r,
```

```
\qquad \qquad \qquad \qquad ''minloss='',round(minloss,6)))
title(sub=paste(''\n'',''\n'',''nlmin.min at
 (a,b)='',round(a[1],4),round(a[2],4),''\n''))
 x_seq(-.5,.5,length=N)
 y_seq(-.5,.5,length=N)
 pd <- persp(x,y,mat2,xlab=''x1'',ylab=''x2'')
 title(sub=paste(''\n'',''\n'',''\n'',''\n'',''\n'',
''Q-optimal density for nu ='',r,''\n''))
list(a = a, minloss = minloss, qwe=qwe)
}
allplots <- function(M=''all'', start0=c(0.05, 0.1), N=101,
width1=3,width2=.2, L=21) {
nu <- c(.2452,.6621,1.2758,2,5,10,100,500,1000,10000)
if(M==''all'') M <- length(nu)
nu <- nu[1:M]
for(k in 1:length(nu)) {
 cat(''\n'',''\n'',''nu='',nu[k],''\n'')
 if(k==1) mystart <- start0
 if(k>1) mystart <- out$qwe$x
 out <- plotloss(r=nu[k], N=N, start=mystart,
\qquad width1=width1,width2=width2, L=L)
 }

invisible()
}

EMP <- function(x,J) {
#This function computes the empirical moment
```

```
#of $X_{1}^{2j}X_{2}^{2k}$

 n <- length(x)/2
 x1 <- x[1:n]
 x2 <- x[(n+1):length(x)]
 X1j <- outer(x1, 2*(1:J), ''^'')
 X2k <- outer(x2, 2*(0:J), ''^'')
 X1k <- outer(x1, 2*(0:J), ''^'')
 X2j <- outer(x2, 2*(1:J), ''^'')
 EMP <- (t(X1j)%*%X2k + t(X2j)%*%X1k)/(2*n)
 for (j in 1:J) { for (k in 1:(J+1))
 { if(k>(j+1)) EMP[j,k] <- 0 }}
EMP
}

expectation_function(j,k,N,mat1){
#This function computes the theoretical moment
#of $X_{1}^{2j}X_{2}^{2k}$

 y1_seq(-.5,.5,length=N)
 y2_seq(-.5,.5,length=N)
 mat2 <- outer(y1^(2*j),y2^(2*(k-1)))
 mat <- mat2*mat1
 simp2(mat)
}

square.error_function(x, J, mat=EXP)sum((EMP(x,J) -mat)^2)
```

```
optim_function(n=6, J=4, N=101) {
#This function matches empirical moment
#with theoretical moment.

 start <- runif(12)/2
 cat(''start='', round(start,3),''\n'')
 MAT1 <- pdensity(N)
 EXP <- matrix(nrow=J, ncol=J+1, data=0)
 for (j in 1:J){ {for (k in 1:(j+1))
 EXP[j,k] <- expectation(j,k,N,mat1=MAT1)
 }}
 cat(''theoretical moments are'',''\n'')
 print.matrix(EXP)
 assign(''EXP'', EXP, frame=1)
 assign(''J'', J, frame=1)
 assign(''n'', n, frame=1)
 fn <- function(x) square.error(x, J=J, mat=EXP)
 qwe <- nlminb(start, fn, lower= rep(0,length(start)),
 upper=rep(.5,length(start)),
 control = nlminb.control(iter.max=500, eval.max = 500,
\qquad x.tol = 1e-10, abs.tol = 1e-10, rel.tol = 1e-10))
 cat(''\n'', ''design points are'', ''\n'')
 print.matrix(round(cbind(qwe$parameters[1:n],
\qquad qwe$parameters[(n+1):12]),3))
 cat(''\n'',''objective ='',qwe$objective,''\n'')
qwe
}
```